

**THE ON-LINE PROCESSING AND  
ANTICIPATION BUILDING IN NATIVE  
MANDARIN SPEAKERS AND LATE DUTCH-  
MANDARIN LEARNERS**

Submitted in partial fulfilment of the requirements

of the degree of

Doctor of Philosophy in Linguistics

PANPAN YAO

SEPTEMBER 2017

SCHOOL OF LANGUAGES, LINGUISTICS AND FILM

QUEEN MARY UNIVERSITY OF LONDON

## **Declaration**

I, Panpan Yao, confirm that the research included within this thesis is my own work or that where it has been carried out in collaboration with, or supported by others, that this is duly acknowledged below and my contribution indicated. Previously published material is also acknowledged below.

I attest that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge break any UK law, infringe any third party's copyright or other Intellectual Property Right, or contain any confidential material.

I accept that the College has the right to use plagiarism detection software to check the electronic version of the thesis.

I confirm that this thesis has not been previously submitted for the award of a degree by this or any other university.

The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without the prior written consent of the author.

Signature:

Date:

## Abstract

In order to explore late L2ers' on-line processing of mass/count syntactic cues which are unique-to-L2 constructions and can only be acquired through implicit learning, a Visual World Paradigm experiment and a Reading for Comprehension experiment were conducted on high proficiency late Dutch-Mandarin learners and native Mandarin speakers.

The mass/count syntactic cues in Mandarin (the Adj-CL word order, and the insertion of the modification maker *de* after a classifier) are unique-to-L2 constructions for Dutch-Mandarin learners, since there exists no classifier in Dutch. Li, Barner, & Huang (2008) is the first experimental study explored native Mandarin speakers' off-line using of the mass/count syntactic cues. They found that nominal phrases with different mass/count syntactic cues led native Mandarin speakers to have different interpretations. It remains unclear how native Mandarin speakers on-line process the mass/count syntactic cues, and whether late L2-Mandarin learners can acquire these mass/count syntactic cues and exhibit native-like behaviours in real time processing. To tackle these questions, the current research tested native Mandarin speakers' and high proficiency late Dutch-Mandarin learners' predictive processing of nominal phrases with different mass/count syntactic cues. Both of the two syntactic cues (the Adj-CL word order, and the insertion of the modification maker *de* after a classifier) were used in the current research, as well as typical count and mass nouns.

The results showed that native Mandarin speakers can take advantage of the mass/count syntactic cues in real time predictive processing. Late Dutch-Mandarin learners exhibited native-like behaviours in the Visual World Paradigm experiment, but non-native-like behaviours in the Reading for Comprehension experiment. The findings indicated that late L2ers can acquire unique-to-L2 constructions through implicit learning, and their processing difficulties are caused by limited cognitive resources but not deficient representations.

## **Dedication**

To my grandma and my parents.

## Acknowledgements

First, I gratefully acknowledge the financial support I have received from the China Scholarship Council (CSC) and Queen Mary University of London during my PhD. Thanks for giving me this opportunity to have a wonderful experience of doing research and living in a big old city.

I would like to express my sincere gratitude to my supervisors Professor Hagit Borer, Doctor Linnaea Stockall, and Doctor David Hall, for their continuous support of my PhD study and related research. Besides my supervisors, I would like to thank Professor David Adger, for being so much more than just a ‘temporary supervisor’. Thank you for all the insightful and inspiring comments about my research.

My sincere gratitude also goes to Professor Lisa Cheng and Professor Rint Sybesma, for their kindly support and help in my participant recruitment and data collection in Leiden University. I also would like to thank Professor Baoguo Chen and Doctor Huanhuan Liu, for their thoughtful help and support during my data collection in Beijing Normal University. It would not have been possible to finish this research project without their help. It is a great pleasure to know these people in my life.

I thank my fellow labmates at the Linguistics Department, Queen Mary University of London, for the great atmosphere and insightful talks and discussions. Also, to all the members of Team China – Annette, Christina, Fangfang, Chen, Orm, Dave, Reem, and Shuting, thanks for being there and comforting me in my bad times, and for all the fun we have had in the last four years.

A special thanks to all of my friends who supported me and companied me in the final stage of my PhD – Reem, Annette, Shuting, Danling, for all the laughter and tears, all the food and drinks, all the encouraging and complaining.

Last but not least, I would like to thank my grandma and my parents, for trusting and supporting me unconditionally throughout my life. Thanks for accepting me to be different.

# Contents

<b>Abstract.....</b>	<b>I</b>
<b>Dedication.....</b>	<b>II</b>
<b>Acknowledgements .....</b>	<b>III</b>
<b>List of Tables .....</b>	<b>VII</b>
<b>List of Figures.....</b>	<b>X</b>
<b>I. Research Questions and Literature Review.....</b>	<b>1</b>
1. Late L2 acquisition and processing of unique-to-L2 constructions .....	4
1.1. The Fundamental Difference Approach .....	5
1.2. The Fundamental Similarity Approach.....	6
1.3. Previous findings .....	7
1.4. Predictions .....	14
2. Anticipation-building in L2 processing.....	16
3. Implicit learning in L2 acquisition .....	22
4. The current research .....	27
<b>II. Classifiers in Mandarin Chinese .....</b>	<b>29</b>
1. English/Dutch vs. Mandarin .....	29
2. Mass/Count in Mandarin Chinese.....	39
3. Mass/count structures are implicit .....	52
<b>III. Previous Studies and Research Methods .....</b>	<b>56</b>
1. Previous studies.....	56
1.1. Classifier studies on native Mandarin speakers.....	57
1.2. Classifier studies on L2-Mandarin learners .....	64
2. Research methods .....	70
2.1. Visual World Paradigm.....	71
2.2. Reading for comprehension task .....	73
2.3. The workload difference .....	78

<b>IV. Visual Word Paradigm Experiment.....</b>	<b>83</b>
1. Material.....	83
1.1. Noun Rating Test: The mass/count categorization of nouns .....	85
1.2. Norming Test.....	91
1.3. Naming Test.....	91
2. Participants.....	92
2.1. Blank-filling Test.....	93
3. Procedure.....	94
4. Prediction.....	94
5. Results .....	99
5.1. Fixation distributions before the nouns .....	99
5.2. Behavioural data and fixation distributions on the nouns .....	110
5.3. The division of the classifiers .....	134
6. Discussion .....	146
6.1. The division of the classifiers .....	147
6.2. Behaviour results .....	149
6.3. Fixation distributions .....	154
<b>V. Reading for Comprehension Experiment.....</b>	<b>164</b>
1. Material.....	165
1.1. Grammaticality Test: The classifier-noun pairs and the sentences .....	167
2. Participants.....	169
3. Procedure.....	169
4. Predictions .....	171
5. Results .....	176
5.1. Processing of the phrases .....	178
5.2. Processing of the classifiers.....	179
5.3. Processing of the nouns .....	182
6. Discussion .....	188
6.1. Classifiers' division on L1ers and L2ers.....	189
6.2. The processing of the classifiers .....	191
6.3. The processing of the nouns .....	195
6.4. The processing of the phrases .....	200
<b>VI. General Discussion .....</b>	<b>205</b>
1. Overview of the current research.....	206

2. L2ers' slowness compared to L1ers' .....	208
3. L2ers' sensitivity to the Adj-Cl word order .....	212
4. L2ers' problems with different classifiers .....	216
5. L2ers' on-line use of Adj information .....	218
6. Conclusions and Contributions .....	219
<b>Appendix A. Materials .....</b>	<b>223</b>
<b>Appendix B. Supplementary Results .....</b>	<b>229</b>
<b>References.....</b>	<b>235</b>



## List of Tables

Table 1. The possibilities of L2ers' general behavioural patterns.....	15
Table 2. The number marking differences among English, Dutch, and Mandarin.....	38
Table 3. The differences between number morphemes and the classifier system .....	39
Table 4. The differences between count-classifiers and massifiers .....	48
Table 5. Different mass/count markings in English/Dutch and Mandarin .....	48
Table 6. A summary of previous studies .....	68
Table 7. Predictions of L1ers' fixation preferences upon hearing each element .....	97
Table 8. Predictions about L2ers' behaviours .....	98
Table 9. Growth Curve Model results in Condition1 in the analysis window of the classifier .....	104
Table 10. Growth Curve Model results in Condition 2 in the analysis window of the classifier and the adjective .....	106
Table 11. Growth Curve Model results in Condition 3 in analysis window of the adjective and the classifier .....	109
Table 12. The best fit model results of behaviour data.....	112
Table 13. The results of Tukey post-hoc test on behavioural data when the classifier is ba.....	112
Table 14. The results of Tukey post-hoc test on behavioural data when the classifier is gen/kuai.....	113
Table 15. L1ers' and L2ers' target picture choices .....	114
Table 16. The cloze probability of count and mass nouns.....	145
Table 17. Interpretation of nominal phrases (%) .....	145
Table 18. The averages of co-occurrence frequency (per million) .....	148
Table 19. The differences between ba and gen/kuai.....	149
Table 20. L1ers' and L2ers' behaviour results.....	150
Table 21. A comparison between predictions vs. L1ers' results ('pre' represents 'prediction', 'res' represents 'results', '=' represents no obvious preference for any picture, CC stands for the count-count picture, CM stands for the count-mass picture, MHA stands for the mass-high-Adj picture, and MLA stands for the mass-low-Adj picture. The check mark $\checkmark$ means L1ers' results are consistent with the predictions. The	

differences between predictions and L1ers' results, and between L1ers' and L2ers' results are highlighted with shadows).....	154
Table 22. A comparison between L1ers' vs. L2ers' results ('L1ers' represents 'native Mandarin speakers' results', 'L2ers' represents 'Dutch-Mandarin learners' results', '=' represents no obvious preference for any picture, CC stands for the count-count picture, CM stands for the count-mass picture, MHA stands for the mass-high-Adj picture, and MLA stands for the mass-low-Adj picture. The check mark √ means L2ers' results are consistent with L1ers. The differences between predictions and L1ers' results, and between L1ers' and L2ers' results are highlighted with shadows.).....	155
Table 23. Examples of the material sentences .....	168
Table 24. The manipulation of the two factors .....	169
Table 25. The cloze probability of count and mass nouns.....	171
Table 26. The interpretation of nominal phrases (%) .....	171
Table 27. The co-occurrence frequency of count nouns and mass nouns (per million) .....	172
Table 28. The differences among the four classifiers .....	172
Table 29. The predictions of processing times on classifiers .....	174
Table 30. The predictions of processing times on nouns.....	174
Table 31. Predictions about L1ers' first fixation durations on classifiers and nouns...	175
Table 32. Predictions about L2ers' behaviours .....	176
Table 33. TRT on the nominal phrase for L1ers.....	178
Table 34. TRT on the nominal phrase for L2ers.....	179
Table 35. The differences of FFD on the classifier for L1ers and L2ers.....	180
Table 36. The differences of FFD on nouns with ba for L1ers.....	183
Table 37. The FFD patterns in the ROI of the nouns for L1ers.....	185
Table 38. The difference of FFD between count and mass nouns with different classifiers for L2ers.....	186
Table 39. The differences of TRT on the nouns .....	187
Table 40. The pattern of processing times in different ROI with different classifiers for L1ers .....	189
Table 41. The differences among the four classifiers .....	190
Table 42. The patterns of processing times in different ROI with different classifiers for L2ers .....	190
Table 43. The predictions of FFD on different classifiers .....	191

Table 44. The prediction of FFD on nouns with different classifiers .....	197
Table 45. A brief summary of the results in the Reading for Comprehension experiment .....	204
Table 46. A summary of main findings .....	220

## List of Figures

Figure 1. Fixation distributions in Condition 1 .....	103
Figure 2. Fixation distributions in Condition 2.....	105
Figure 3. Fixation distributions in Condition 3.....	107
Figure 4. Analysis window of the noun in Condition 1 with ba .....	117
Figure 5. Analysis window of the noun in Condition 2 with ba .....	119
Figure 6. Analysis window of the noun in Condition 3 with ba .....	121
Figure 7. Analysis window of the noun in Condition 1 with gen .....	123
Figure 8. Analysis window of the noun in Condition 2 with gen .....	125
Figure 9. Analysis window of the noun in Condition 3 with gen .....	127
Figure 10. Analysis window of the noun in Condition 1 with kuai .....	129
Figure 11. Analysis window of the noun in Condition 2 with kuai .....	131
Figure 12. Analysis window of the noun in Condition 3 with kuai .....	132
Figure 13. The analysis window of the classifier in Condition 1 .....	135
Figure 14. Analysis window of the classifier in Condition 2 .....	137
Figure 15. Analysis window of the classifier in Condition 3 .....	139
Figure 16. First fixation duration on classifiers .....	181
Figure 17. Total reading times on classifiers .....	182
Figure 18. First fixation duration on nouns with different classifiers for L1ers.....	185
Figure 19. First fixation durations on the nouns .....	186
Figure 20. Total reading times on the nouns.....	187

## I. Research Questions and Literature Review

The concern of the current research is second language processing (L2 processing) of late second language learners (late L2ers), focusing on syntactic effects on real-time predictive processing. The research questions are whether late L2ers can eventually exhibit native-like behaviours during on-line predictive processing of unique-to-L2 constructions, and if not, whether their processing difficulties are caused by a deficit in L2 representations or a deficit in performance.

Most researchers agreed that late L2ers have some processing difficulties when compared to native speakers (see a review of Slabakova, 2016). However, it is still debatable whether the observed L2 processing difficulties are caused by a deficit in L2 representation or a deficit in performance. Also, it still remains unclear whether late L2ers can eventually acquire native-like L2 representations and exhibit native-like behavioural patterns. The Representation Deficit Account (Bley-Vroman, 1989, 2009; DeKeyser, 2000; Hawkins and Hattori, 2006; Tsimpli and Dimitrakopoulou, 2007; Schachter, 1998) claims that compared to native speakers, late L2ers have deficient L2 representations even at the end-state of L2 acquisition. This is due to their late acquisition of L2 after puberty. Their observed L2 processing difficulties are caused by the deficit in their representations. On the other hand, the Performance Deficit Account (Dekydtspotter, Schwartz and Sprouse, 1996; Hopp, 2006; White, 2003) claims that the L2 processing difficulties are caused by L2ers' performance (e.g., limited cognitive resources) instead of their deficient representations. With increased L2 proficiency and low processing load, late L2ers can exhibit native-like patterns in real time processing.

Among several aspects affecting late L2ers' processing, first language (L1) transfer has attracted a lot of research interest (Ionin, et al., 2010, 2012, 2013; Jiang, 2004, 2007; Jiang, Novokshanova, Masuda and Wang, 2011; Slabakova, 2008). The role of L1 in L2 acquisition and processing has been much debated and discussed. The issue is no longer 'whether L1 has an effect on L2 acquisition and processing' since most researchers have achieved an agreement that at least some aspects of L2 acquisition and processing are

influenced by L2ers' native language (N.C. Ellis, 2006; Gass & Selinker, 1992; Schwartz & Sprouse, 1996). Instead, what remains unclear is whether it is possible for late L2ers to recover from L1 transfer and acquire unique-to-L2 knowledge which is not instantiated in their L1. Based on the Morphological Congruency Hypothesis (Jiang et al., 2011), late L2ers' representations are constrained by their L1: only congruent morphological structures which exist in both L1 and L2 in a similar way can be acquired; unique-to-L2 knowledge is very difficult, if possible at all, to acquire for late L2ers. On the other hand, the Feature Reassembly Hypothesis (Lardiere, 2009) claims that there is no permanent effect of L1 in L2 acquisition. Even though L2ers transfer L1 grammar properties in their L2 processing at the initial stage, they are able to recover from this L1 transfer and reassemble new features in L2 grammar which are motivated by the evidence of L2 input.

Based on previous studies in the last two decades, in the current study, I argue that even though they start to learn L2 after puberty, late L2ers can ultimately acquire native-like L2 representations. Their processing difficulties are caused by a deficit in performance (i.e. limited cognitive resources and L2 proficiency), but not by deficient representations. Also, with increased L2 proficiency, late L2ers are able to recover from L1 transfer and acquire unique-to-L2 knowledge.

In this current research, two eye-tracking experiments with different workloads (a Visual World Paradigm experiment with a comparatively low workload and a Reading for Comprehension experiment with a comparatively high workload)<sup>1</sup> were conducted on high proficiency late Dutch-Mandarin learners<sup>2</sup> and native Mandarin speakers. The mass/count syntax marked in the classifier system in Mandarin was used as an implicit and unique-to-L2 construction<sup>3</sup> for Dutch-Mandarin learners. Different from English and Dutch, in which the mass/count distinction is marked through plural morphemes and

---

<sup>1</sup> Detailed discussions about the difference of workload requirements between the Visual World Paradigm experiment and the Reading for comprehension experiment are presented in Chapter 3.

<sup>2</sup> Adopting the term from previous studies in psychology and psycholinguistics fields, in this thesis 'Dutch-Mandarin learners' is used to refer to Dutch speakers who learn Mandarin as their second language. With the term 'Dutch-Mandarin learners', the language before the dash refers to the native language, while the language after the dash refers to the second language and also the target language in the current research.

<sup>3</sup> The implicitness and uniqueness of the materials used in the current research are discussed and validated in Chapter 2.

articles, in Mandarin, the mass/count distinction is encoded in the classifier system according to Cheng & Sybesma (1998, 1999). The mass/count structures<sup>4</sup> in Mandarin are thus unique-to-L2 knowledge to Dutch-Mandarin learners. Comparing high proficiency late Dutch-Mandarin learners' on-line processing of this unique-to-Mandarin construction to native Mandarin speakers' behaviours, and comparing late Dutch-Mandarin learners' performance between the two eye-tracking experiments with low and high workloads (the Visual World Paradigm experiment and the Reading for Comprehension experiment), one can tell whether late L2ers can eventually acquire unique-to-L2 knowledge and exhibit native-like behaviours during on-line processing (the Representation Deficit Account vs. the Presentation Deficit Account; the Morphological Congruence Hypothesis vs. the Feature Reassembly Hypothesis).

The results show that high proficiency late Dutch-Mandarin learners exhibited native-like behavioural patterns when using mass/count syntactic cues during on-line processing, which indicates that late L2ers can eventually acquire native-like L2 representations and exhibit native-like behaviours during on-line processing of unique-to-L2 constructions. Late L2ers can ultimately acquire unique-to-L2 constructions, and their processing difficulties are due to their performance deficit but not representation deficit. This finding supports my argument, the Performance Deficit Account (Dekydtspotter, Schwartz and Sprouse, 1996; Hopp, 2006; White, 2003) and the Feature Reassembly Hypothesis (Lardiere, 2009).

This thesis is structured as follows. In Chapter 1 existing theories and observations about L2 acquisition and processing are analysed, based on which the research questions are raised. In Chapter 2 the main differences between the grammaticalized mass/count distinction in Dutch and Mandarin are introduced, and the uniqueness and the implicitness of the mass/count structures in Mandarin nominal phrases for Dutch-Mandarin learners are proposed. In Chapter 3, the research methods employed in the current research to look into participants' on-line processing and implicit learning are introduced and discussed, along with previous research focused on the acquisition of

---

<sup>4</sup> In this thesis, the term 'mass/count structure' is used to refer to nominal phrases with different mass/count syntactic cues, i.e. nominal phrases with the same morphological content but different word orders. Introductions and discussions about the 'mass/count structure' are presented in Chapter 2. Detailed information of the materials used in each experiment is presented in Chapter 4 and Chapter 5.

mass/count semantics and syntax, and the classifier system in Chinese. Detailed information of the two main experiments in the current research (the Visual World Paradigm experiment and the Reading for Comprehension experiment) is presented in Chapter 4 and Chapter 5 respectively. In Chapter 6 a general discussion is presented, with some suggestions outlined for future research.

Language processing is a very complicated procedure, containing different sub-procedures and the interactions among them. In this current research, the critical research interest is about late L2ers' real-time anticipation building using unique-to-L2 constructions, which can only be learned through implicit learning. To have a clear understanding of the background of this language processing procedure, I break it into three sub-parts: late L2 acquisition and processing of unique-to-L2 constructions, L2 predictive processing in real time, and implicit learning in L2 acquisition. Relevant theories and previous studies in these three specific sub-parts are introduced individually in this chapter.

### **1. Late L2 acquisition and processing of unique-to-L2 constructions**

The Critical Period Hypothesis (Lenneberg, 1967) states that due to maturational reasons, language acquisition is constrained by language-learners' age. If language acquisition does not occur by puberty, some aspects of language (e.g. grammar, morphosyntax) cannot be fully mastered by language learners. Extended to the Second Language Acquisition (SLA) field, the Critical Period Hypothesis holds that after puberty L2 can no longer be acquired by the innate language acquisition device and as implicit and proceduralized knowledge, but must be learned incompletely through explicit instructions and long-time practice (Krashen, 1994; MacWhinney, 2005). Based on the existing findings over the last 30 years, researchers agree that if the age of acquisition is after puberty, most L2ers exhibit some processing difficulties (such as, slowness, low accuracy, non-native-like patterns) in L2 processing (see a review of Slabakova, 2016). The issue still under debate is whether late L2ers' processing difficulties are caused by their representation deficits or performance deficits. And whether L1 has a permanent effect in L2 acquisition.



In this section, the relevant theories in late L2 acquisition and processing are introduced and discussed first, with respect to both sides of the division – the Fundamental Difference Approach (both the Representation Deficit Account and the Morphological Congruency Hypothesis argue for a fundamental difference between L2ers' representations and native speakers') and the Fundamental Similarity Approach (both the Performance Deficit Account and the Feature Reassembly Hypothesis claim that when reach an advanced proficiency level, there is no fundamental representational difference between L2ers and native speakers). Following which, existing findings of late L2ers' acquisition and processing are discussed in relation to these theories. Based on these theories and previous studies, my arguments and assumptions are raised.

### **1.1. The Fundamental Difference Approach**

The Representation Deficit Account arguing for a fundamental difference between native speakers and L2ers claims that the maturational constraints (around puberty) result in a loss of plasticity in late L2ers' brains, and as a consequence, they cannot acquire L2 in the same way as native speakers do, and need to resort to a different learning mechanism and different linguistic representations. Thus late L2ers' grammatical representations and language processing routines are fundamentally different from those of native speakers. They cannot exhibit native-like behaviours since they do not acquire native-like L2 representations (Bley-Vroman, 1989, 2009; DeKeyser, 2000; Hawkins and Hattori, 2006; Tsimpli and Dimitrakopoulou, 2007; Schachter, 1998).

Regarding the L1 transfer effect on late L2ers' acquisition and processing, Jiang et al. (2011) propose the Morphological Congruency Hypothesis. They argue that languages vary in terms of which meaning is grammaticalized and morphologically marked. The language distance/cross-linguistic relationship can be divided into two levels of morphological congruency based on morphological marking: languages that are morphologically congruent and languages that are morphologically incongruent. In the former a certain meaning is morphologically grammaticalized in both L1 and L2, while in the latter a grammatical morpheme is present only in one of the two languages. It should be noted that morphological congruency refers specifically to a particular morpheme, rather than two languages in general. For example, French and English are morphologically congruent in plural marking but morphologically incongruent in gender marking. In order to learn an L2, L2ers need to learn every grammatical morpheme that does or does not have a counterpart in their L1s, thus the morphological congruency

between their L1 and L2 affects their ultimate attainment in L2: only the congruent L2 knowledge (the grammatical morphemes which exist in both L1 and L2) can be fully acquired by late L2ers when reaching a near-native level. While for the incongruent L2 knowledge (the grammatical morphemes which exist only in L2 but not in L1) it is extremely difficult, if possible, to develop a native-like representation.

To sum up, the Representation Deficit Account argues that late L2ers have different representations from native speakers, and their processing difficulties and non-native-like patterns are caused by these deficient representations. The Morphological Congruency Hypothesis argues that only congruent knowledge can be acquired by late L2ers. It is very difficult for late L2ers to exhibit native-like patterns in processing of unique-to-L2 knowledge. To further extend, both the Representation Deficit Account and the Morphological Congruency Hypothesis predict that late L2ers' processing may improve with increased L2 proficiency, but the end-state late L2ers (late L2ers who have reached a near-native L2 proficiency level) would still display non-native-like behaviours since their L2 representations are fundamentally different from those of native speakers, especially when they processing unique-to-L2 constructions.

## **1.2. The Fundamental Similarity Approach**

The Performance Deficit Account arguing for fundamental similarity between native speakers and L2ers claims that late L2ers' grammatical representations and language processing routines are quantitatively, but not qualitatively different from native speakers'. Late L2ers can eventually acquire native-like L2 representations; their L2 processing is limited by their cognitive resources and proficiency levels (Dekydtspotter, Schwartz and Sprouse, 1996; Hopp, 2006; White, 2003). Compared to native speakers, L2ers have two possible linguistic routes in their mind when processing sentences in real-time: their L1 and L2. Increased processing demand is caused by having to identify words, phrases, and grammatical content in L2. As a consequence, L2 on-line processing requires more cognitive resources or imposes a higher workload compared to the same procedure in L1. However, the command of L2 grammar can be improved through long-term practice and increasing L2 proficiency. Thus, as long as they reach a very high proficiency level of L2, late L2ers should be able to exhibit native-like behaviour patterns in real-time sentence processing.

Regarding the L1 transfer effect on late L2ers' acquisition and processing, Lardiere (2009) proposes the Feature Reassembly Hypothesis. Different from the Morphological

Congruency Hypothesis which predict a permanently constraining role of L1 in L2 acquisition, the Feature Reassembly Hypothesis argues that the restriction from L1 is limited and can be overcome with the increase of L2 proficiency. At the initial stage of L2 acquisition, L1 feature values may transfer and affect late L2ers' processing of L2. In other words, L2ers map L1 grammars into L2 processing. However, when L2ers have reached an advanced level of L2 proficiency through immersive L2-dominant environment, loads of L2 input and positive evidence enable them to reassemble new feature values and eventually restructure their L2 grammars accordingly. In this case, some incongruent and unique-to-L2 constructions can be ultimately acquired by late L2ers through this 'feature reassembly' procedure. According to Lardiere (2009), there are two main learning tasks for L2ers in L2 acquisition: mapping and reassembly. First of all, they need to 'identify one or more lexical items over which to redistribute the features associated with a particular functional element in the L2' (Lardiere 2009, 174). And after that, they need to 'acquire new language-specific configurations of features as these are assembled in the targeted lexical items of the L2' (Lardiere 2009, 175).

To sum up, the Performance Deficit Account argues that late L2ers have no qualitative difference from native speakers, they can eventually acquire native-like representations and exhibit native-like patterns after long-term practice and with increased L2 proficiency. Late L2ers' processing difficulties are restricted by their proficiency levels and limited cognitive resources. The Feature Reassembly Hypothesis argues that even though late L2ers map their L1 grammar in L2 processing at the initial stages, they would eventually recover from the L1 transfer and reassemble new features in L2 grammar. Both the Performance Deficit Account and the Feature Reassembly Hypothesis share the prediction that with increased L2 proficiency, and/or a low workload, late L2ers can display native-like behaviours, overcome the L1 transfer effect and acquire native-like L2 representations even if some of the L2 constructions only exist in L2.

### **1.3. Previous findings**

Second language (L2) acquisition is a challenging task especially for late L2ers (Chen, Shu, Liu, Zhao & Li, 2007; Dekeyser, 2005; Hahne & Friederici, 2001; Jiang 2004, 2007; Lew-Williams & Fernald, 2010; Ojima, Nakata & Kakigi, 2005). It is fairly well established that if L2 acquisition happens after puberty, it is difficult for L2ers to have exactly native-like behaviour patterns. Much research has revealed a variety of non-native-like behaviours in late L2ers. First of all, L2ers are generally slower than native

speakers (L1ers) in sentence processing in both on-line and off-line tasks, as demonstrated across a variety of tasks and linguistic structures (Hahne & Friederici, 2001; Lew-Williams & Fernald, 2010; Sanders & Neville, 2003; Trenkic, Mirkovic, & Altmann, 2013). Secondly, late L2ers exhibited better performance in off-line tasks than on-line tasks (for review, see Clahsen & Felser, 2006). Moreover, late L2ers' proficiency has been found to have an influence on L2 processing (Clahsen & Felser, 2006; French-Mestre, 2002; Ionin and Montrul, 2010; Perani et al., 1998; Slabakova, 2016). Late L2ers with different L2 proficiency levels exhibited different processing patterns: with increased L2 proficiency level, it is increasingly possible for late L2ers to exhibit native-like performance in behavioural experiments (Ionin and Montrul, 2010; Leal, et al., 2016; Hopp, 2006), eye-tracking studies (French-Mestre, 2002) and Event-Related Potential (ERP) studies (Hahne, 2001; Hahne and Friederici, 2001; Sabourin, 2003). Furthermore, different linguistic modules evoke different behaviours in late L2ers. Some studies found that different from native speakers who can take advantage of both lexical-semantic information and syntactic information during parsing and production, late L2ers can only process lexical-semantic information in a native-like manner, but not complex syntactic information (Clahsen & Felser, 2006; Lau & Gruter, 2015; Lew-Williams & Fernald, 2009; Felser, Roberts, Marinis, & Gross, 2003). Late L2ers often show some difficulties and non-native-like behaviours in automatically applying morphology, complex syntax, and discourse pragmatics (Benmamoun, Montrul, and Polinsky, 2013a, 2013b; Hahne, Mueller, & Clahsen, 2006; Kaan, 2014; Parodi, Schwartz & Clahsen, 2004). In spite of these findings, there exist inconsistent observations about late L2ers' behaviour in the processing of unique-to-L2 constructions. Some studies found that late L2ers exhibit non-native-like behaviour patterns with some 'unique-to-L2' structures (Chen et al., 2007; Ionin et al., 2012; Jiang, 2004, 2007; Kotz, 2009). On the other hand, native-like behaviours in late L2ers have been found with unique-to-L2 constructions in behavioural experiments (Hopp, 2006; Ionin and Montrul, 2010; Jackson, 2007; Jackson & Dussias, 2009), eye-tracking studies (Yao & Chen, 2016), and ERP studies (French-Mestre, 2009; Morgan-Short et al., 2010). It is still unclear based on these existing findings whether late L2ers can eventually exhibit native-like behaviours in the processing of unique-to-L2 constructions.

Taken together, late L2ers' behaviour in L2 acquisition and processing can be summarized as follows:

- In general, late L2ers are slower than native speakers
- Late L2ers exhibit better performance in off-line tasks than in on-line tasks
- Increased L2 proficiency brings more native-like behaviours in late L2ers
- Compared to lexical-semantic information, L2ers under-use complex syntactic/morphological information in on-line tasks
- Native-like and non-native-like behaviour can be found in processing of unique-to-L2 constructions

These existing findings can be explained by one or both of the approaches regarding late L2 acquisition and processing (the Fundamental Difference Approach vs. the Fundamental Similarity Approach).

First of all, the finding that late L2ers are slower than native speakers in general language processing can be explained by both the Representation Deficit Account and the Performance Deficit Account. From the representation deficit view, late L2ers have deficient L2 representations compared to native speakers due to their late acquisition after puberty. These deficient L2 representations lead to late L2ers' slowness in processing L2 compared to native speakers. On the other hand, from the performance deficit view, late L2ers have native-like L2 representations, but their performance is restricted by their limited cognitive resources and/or L2 proficiency due to the competition between L1 and L2 in real-time processing. Late L2ers' processing slowness is caused by their imperfect performance. Once they have reached an advanced level of L2 proficiency, or in a low-workload-demanded task, late L2ers are able to exhibit native-like performance.

The finding that late L2ers exhibit better performance in off-line tasks than in on-line tasks can be explained by both the Representation Deficit Account and the Performance Deficit Account. Compare to off-line tasks, on-line tasks require heavier time restrictions. According to the Representation Deficit Account, in off-line tasks where there is no time limit, late L2ers can take their time to process deficient L2 representations and use some explicit processing strategies to compensate for their lack of 'full' representations. In this case, late L2ers can exhibit indistinguishable behaviours from native speakers in off-line tasks. While in on-line tasks they perform differently from native speakers since the limited time leads to their incomplete processing of deficient representations and

inability to access explicit processing strategies. Thus it is reasonable to expect late L2ers exhibit better performance in off-line tasks than in on-line tasks. On the other side, late L2ers' better performance in off-line tasks than in on-line tasks can also be explained by the Performance Deficit Account: their processing difficulties are caused by their limited cognitive resources and L2 proficiency. Late L2ers' native-like performance in off-line tasks can be treated as the evidence that they have acquired native-like L2 representations. Their performance difficulties in on-line tasks are caused by the heavy workload of on-line tasks and their limited cognitive resources.

Before we discuss the effect of proficiency in L2 processing, it is necessary to figure out what L2 proficiency is. Is L2 proficiency an index of L2 representation, or L2 performance, or both, or neither? The majority of previous studies explored the effect of L2 proficiency in L2 processing without giving it a clear and specific definition. Considering that L2 proficiency is not a manipulated and comparable factor in the current study (I did not manipulate the L2 proficiency to have different levels and compare L2 participants' performance directly among different L2 proficiency levels), I decide to go along with the previous studies (Dussias et al., 2013; Kaan, 2014; Leal et al., 2016) in defining it as a learner's knowledge of L2 structures and the skill in using it. In other words, in this current research, L2 proficiency refers to both L2 representations and L2 performance.

In this sense, the proficiency effect in L2 processing can be explained by both approaches. According to the Representation Deficit Account and the Morphological Congruency Hypothesis, at the initial stage of L2 acquisition, all the start points of L2 grammar are built based on L2ers' L1 knowledge. After that, as they become more and more proficient in L2, more and more representations are becoming native-like, even though some L2 knowledge (especially those uninterpretable unique-to-L2 knowledge) can never be acquired in a native-like way. Thus with increased L2 proficiency, late L2ers are expected to exhibit more native-like behaviours. On the other hand, both the Performance Deficit Account and the Feature Reassembly Hypothesis also predict that increased L2 proficiency will lead to more native-like patterns. To be specific, with increased L2 proficiency, more and more L2 grammars have been acquired by late L2ers through their 'feature reassembly' procedure, and their ability of automatically using these features has also been improved. Thus increased L2 proficiency will definitely lead to more native-like patterns.

Regarding late L2ers' better performances on lexical-semantic information than complex syntactic and morphological information, both the Representation Deficit Account and the Performance Deficit Account can explain this phenomenon. The Representation Deficit Account proposes that compared to native speakers, late L2ers have deficient L2 representations. During L2 processing, late L2ers rely more on surface knowledge (e.g. lexical, semantic knowledge) to compensate for the deficient representation of complex syntactic knowledge. Following this logic, it is reasonable to observe late L2ers' better performance on lexical-semantic information than complex syntactic and morphological information. On the other hand, according to the Performance Deficit Account, late L2ers' processing is restricted by their limited cognitive resources. Considering the lexical information is more 'superficial' than syntactic information, which needs less cognitive resources than the latter one when being processed, it is reasonable for late L2ers to have better performance on lexical information than complex syntactic information.

So far, both the Fundamental Difference Approach and the Fundamental Similarity Approach can explain most of the existing findings in late L2 processing. However, when it comes to the variable observations in relation to late L2ers' processing of unique-to-L2 constructions, the Fundamental Similarity Approach, especially the Feature Reassembly Hypothesis has a better explanation.

With respect to the variable findings of late L2ers' processing of unique-to-L2 constructions, it is very complicated since there are many different factors which may affect experimental results and the corresponding interpretations. For example, research methods (on-line vs. off-line, high workload vs. low workload), language pairs of L2ers (L1-L2 distance is far or close), the definitions and manipulations of L2 proficiency (proficiency test, self-report, length of L2 learning, duration of living experience in a L2 dominant environment), and the testing language environment (in L1 or L2 environment, with L1 or L2 instructions). Taking a closer look at previous studies, I found that different research methods affect late L2ers' performance: more non-native-like behaviours were observed in on-line processing tasks (Jiang, 2004, 2007, 2011; Chen et al., 2007) than in off-line processing tasks (Ionin et al, 2012). Also, tasks with different workloads affect late L2ers' performance: more native-like behaviours were found in tasks with low workloads than in high-workload-tasks (Foucart & Frenck-Mestre, 2012; Hopp, 2006; Yao & Chen, 2016). This finding indicates that the workloads of tasks affect

late L2ers' performance, which is consistent with the prediction of the Performance Deficit Account that late L2ers' limited cognitive resources restrict their performance.

Moreover, late L2ers' L1-L2 distance affects their performance: more native-like behaviours were found with close L1-L2 distance than with far L1-L2 distance (Hopp, 2006; Jiang et al., 2011; Ionin & Montrul, 2010). This is consistent with the Performance Deficit Account that when the L1-L2 distance is far, the L1-L2 competition needs more cognitive resources than when the L1-L2 distance is close. Since late L2ers' performance is limited by their cognitive resources, it is reasonable to observe more native-like patterns with close L1-L2 distance (low cognitive resource required) than with far L1-L2 distance (high cognitive resource required).

Different definitions and manipulations of L2 proficiency lead to different interpretations of experimental results: some studies used late L2ers' scores in self-report and some language tests to be the index of their L2 proficiency (Chen et al., 2007; Jackson & Dussias, 2009; Yao & Chen, 2016), while other studies used late L2ers' length of L2 exposure and duration of L2 learning (Hopp, 2006; Ionin et al., 2013; Jiang, 2011). The standard of high or low proficiency is not consistent. Thus 'high proficiency L2 participants' in some studies may be 'intermediate proficiency L2 participants' in others, and vice versa. As a consequence, researchers have different interpretations of their research findings: some claim that late L2ers can eventually acquire L2 representation and exhibit native-like behaviours (Foucart & Frenck-Mentre, 2012; Hopp, 2006; Ionin et al., 2010, 2013; Jackson & Dussias, 2009; Yao & Chen, 2016), while others claim that even advanced L2ers cannot behave like native speakers (Ionin et al., 2012; Jiang, 2004, 2007, 2011). Previous studies illustrate that more native-like behaviours were observed on high to advanced L2ers when their L2 proficiency was mainly judged by their length of L2 exposure (Foucart & Frenck-Mentre, 2012; Hopp, 2006; Ionin et al., 2013) than when their L2 proficiency was judged by the scores of self-report and language tests (Jackson & Dussias, 2009; Yao & Chen, 2016). This finding indicates that compared to the scores in language tests and self-reports, the length of L2 exposure may be a better index for late L2ers' proficiency level. Additionally, with increased length of living experience in a L2-dominant environment, more native-like patterns were observed. This is consistent with the prediction of the Performance Deficit Account that late L2ers can eventually exhibit native-like behaviours when they reach a near-native level, since they do not have deficient representations.



The language testing environment seems to have no effect on late L2ers' performance: native-like behaviours were spotted in both L1-dominant environment (Ionin et al., 2010; Jackson & Dissias, 2009; Yao & Chen, 2016) and L2-dominant environment with L2 instructions (Foucart & Frenck-Mestre, 2012; Hopp, 2006; Ionin et al., 2013).

From these analyses we can see that, it is increasingly possible to observe late L2ers' native-like behaviours, 1) when the research method requires low processing loads (off-line tasks and low-workload paradigms); 2) when the L1-L2 competition needs low cognitive resources (close L1-L2 distance); 3) when late L2ers have a long period of living experience in a L2-dominant environment. These findings are consistent with the predictions based on Performance Deficit Account and the Feature Reassembly Hypothesis, which claim that late L2ers can ultimately acquire unique-to-L2 constructions; their processing difficulties are caused by their limited cognitive resources and L2 proficiency. Between the Fundamental Difference Approach and the Fundamental Similarity Approach, the latter offers a better explanation of late L2ers' processing of unique-to-L2 constructions.

So far, most of the existing observations can be explained by both the Fundamental Difference Approach and the Fundamental Similarity Approach. While regarding late L2ers' varied behaviours in the processing of unique-to-L2 constructions, the Performance Deficit Account and the Feature Reassembly Hypothesis offer a better explanation.

Based on these findings from previous studies, I argue that late L2ers' performance is more restricted by their limited cognitive resources and L2 proficiency than their L2 representations. There is no fundamental difference between native speakers' and late L2ers' representations. Late L2ers' observed L2 processing difficulties are caused by their limited cognitive resources and L2 proficiency, but not their deficient representations (the Performance Deficit Account). When they reach an advanced L2 proficiency level, it is highly possible for them to acquire and process unique-to-L2 constructions in a native-like manner (the Feature Reassembly Hypothesis).

The current research is set up to test this hypothesis that late L2ers can eventually acquire unique-to-L2 constructions and behave in a native-like way during on-line processing, they will exhibit better performance in the task with a low workload than in the task with a high workload.

#### 1.4. Predictions

From Section 1.1 and Section 1.2, we can see that the critical difference between the two approaches is whether the end-state L2ers can acquire native-like L2 representations and exhibit native-like behaviours during on-line processing of unique-to-L2 constructions, and whether tasks with different workloads (high vs. low) will affect L2ers' performance. The Representation Deficit Account predicts that end-state L2ers cannot acquire native-like L2 representations and cannot exhibit native-like behaviours. Increased L2 proficiency will not overcome the processing difficulties caused by the deficient L2 representations. Moreover, tasks with different workloads will not affect high proficiency L2ers' performance since the problem is at the representation level but not the performance level. On the other hand, the Performance Deficit Account predicts that end-state L2ers can acquire native-like L2 representations, and exhibit native-like behaviours during on-line processing. Their deficit is at the performance level but not the representation level. Thus with decreased workloads and increased L2 proficiency, it is possible to observe late L2ers' native-like behaviours.

In the current research, along with the Performance Deficit Account and the Feature Reassembly Hypothesis, I argue that even after puberty, late L2ers can eventually acquire native-like L2 representations. Their processing difficulties are not caused by their deficient representations, but their limited cognitive resources and L2 proficiency. They can acquire and exhibit native-like performance in on-line processing of unique-to-L2 constructions.

To investigate whether late L2ers can eventually exhibit native-like behaviours in on-line processing of unique-to-L2 constructions, and whether tasks with different workloads affect L2ers' performance, in the current research, two eye-tracking experiments were conducted on highly proficient Dutch-Mandarin learners. Compared to the Reading for Comprehension experiment, the Visual World Paradigm experiment requires a lower workload on participants. Comparing high proficiency late Dutch-Mandarin learners' performance between these two eye-tracking experiments, there are four possibilities of their general behavioural patterns, which are summarized in Table 1. Participants' possible behaviour patterns in the two experiments are listed in the first two columns, and the theories each possible behavioural pattern would support are listed in the third column.

Table 1. The possibilities of L2ers' general behavioural patterns

<b>Visual World Paradigm (low workload)</b>	<b>Reading for Comprehension (high workload)</b>	<b>Support theories</b>
Native-like	Native-like	Performance Deficit Account & Feature Reassembly Hypothesis
Native-like	Non-native-like	Performance Deficit Account & Feature Reassembly Hypothesis
Non-native-like	Native-like	Not possible
Non-native-like	Non-native-like	<ul style="list-style-type: none"> <li>•Representation Deficit Account &amp; Morphological Congruency Hypothesis</li> <li>•Performance Deficit Account &amp; Feature Reassembly Hypothesis</li> </ul>

To be specific, in the current research, if high proficiency late Dutch-Mandarin learners exhibit native-like patterns in both of the two experiments, the results would support the Performance Deficit Account and the Feature Reassembly Hypothesis, since this native-like pattern in both experiments can be treated as the evidence that they have already acquired the unique-to-L2 constructions. The second possibility is Dutch-Mandarin learners only exhibit native-like patterns in the Visual World Paradigm which has a lower workload compared to the Reading for Comprehension Task. This result would indicate that they have acquired the unique-to-L2 constructions, but their performance is limited by the processing loads of tasks. In this case, these results will also support the Performance Deficit Account and the Feature Reassembly Hypothesis. The third possibility that they exhibit non-native-like patterns in the Visual World Paradigm but native-like patterns in the Reading for Comprehension Task is considered to be not reasonable and practical in the current research, since I argue that participants should always have better performance in a low-workload-demanding task (the Visual World Paradigm) than in a high-workload-demanding task (the Reading for Comprehension Task). Finally, if they exhibit non-native-like patterns in both of the experiments, there could be two ways to interpret this result. One is treating their non-native-like patterns as the evidence that they can never acquire native-like representations of unique-to-L2 constructions. In this case, the results will support the Representation Deficit Account

and the Morphological Congruency Hypothesis. On the other hand, if we consider that the Dutch-Mandarin learners in the current study were not end-state L2ers (they were at their high proficiency level of Mandarin back then), it is possible to argue that the Dutch-Mandarin participants were still in the process of acquisition, it would still be possible for them to ultimately acquire the native-like representations when reach the end-state level.

Based on the Performance Deficit Account and the Feature Reassembly Hypothesis, I predict that, high proficient late Dutch-Mandarin learners will exhibit native-like performance during on-line processing, and it is more probable to observe their native-like behaviours in the Visual World Paradigm experiment (with a comparatively low workload) than in the Reading for Comprehension experiment (with a comparatively high workload).

## **2. Anticipation-building in L2 processing**

During sentence comprehension, people need to integrate different sources of information (e.g. lexical, semantic, syntactic, etc.) rapidly, and combine these different sources of input to build an evolving representation and possible interpretations. Previous studies in real time sentence processing found that native speakers' sentence comprehension is not a passive integration of different types of linguistic input, but a proactive procedure of building anticipations and constructions based on the input. Anticipation building is a key part of on-line sentence processing and comprehension (Altmann & Mirkovic, 2009; Federmeier, 2007; Kamide 2008; Levy, 2008; Pickering & Garrod, 2013). Huettig, Rommers & Meyer (2011) found that native speakers can take advantage of different types of linguistic input to rapidly make predictions in real time processing. More studies focusing on native Chinese speakers' anticipation building during on-line processing are presented in Chapter 3.

In the field of L2 processing, findings of late L2ers' real-time prediction-building are not conclusive. Some studies found that, compared to native speakers, L2ers exhibit a reduced ability in using the current information to anticipate upcoming items (Lew-Williams & Fernald, 2010; Grüter, Lew-Williams & Fernald, 2012). Kaan, Dallas & Wijnen (2010) proposed that late L2ers' difficulties in morphosyntactic processing are

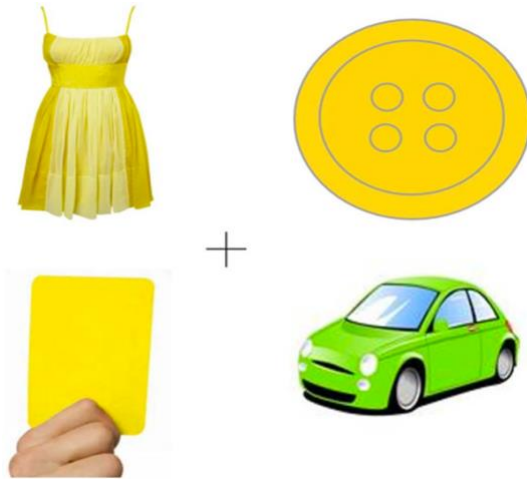
associated with their limited ability in using syntactic information to make predictions in on-line processing. Further, Grüter & Rohde (2013) argued that L2ers' reduced ability to make predictions in real-time processing is a general ability deficit, and that they would exhibit anticipatory problems no matter whether the cue information is lexical, semantic, or morphosyntactic. These studies indicated a prediction difficulty in real-time L2 processing. On the contrary, Köhne & Crocker (2010) found that, similar to native speakers, L2ers can take advantage of lexical information and context cues to anticipate upcoming referents in real-time processing. Foucart, Martin, Moreno & Costa (2014) showed that L2ers can use high semantic constraining context information to predict upcoming words. These studies illustrate that L2ers can use the semantic and context information to form anticipations during on-line processing.

One of the main concerns of the current research is whether late L2ers can take use of unique-to-L2 knowledge to form anticipations during on-line processing. In this section, recent studies looking into late L2ers' predictive use of unique-to-L2 knowledge in real-time are summarised and analysed.

Hopp (2012) used a Visual World Paradigm experiment to test the predictive processing of syntactic gender agreement between determiners and nouns in English-German learners and native German speakers. Twenty advanced to near-native English-German learners were recruited as L2 participants. Gender marking only exists in German but not in English, thus it can be treated as unique-to-L2 knowledge to English-German learners. Based on the Representation Deficit Account he argued that if late L2ers have some deficits in their L2 knowledge representations, they would initially pair all nouns in the L2 with the most frequently occurring determiner as a default gender instead of the appropriate gender determiner. The gender status of the determiner in the audio sentence was manipulated to have three levels: masculine, feminine, and neuter. An ambiguous adjective which can occur with all three gender levels of nouns were used. The example of the materials is illustrated in (1).

(1) Wo ist der/die/das gelbe [Noun]?

Where is the<sub>MASC/FEM/NEUT</sub> yellow [Noun]?



Hopp (2012)

Participants were required to look at the four pictures presented on the screen, and at the same time listen to the instructions from the headphones, and respond as the instructions indicated. Focusing on which picture participants start to land more of their fixation on when hearing the gender-marked determiner, the author could tell whether participants are using the gender information from the determiner to predict the upcoming gender-consistent noun. The results yield differences in anticipatory use of the gender information of the determiner between the advanced L2 group and the near-native L2 group: the near-native English-German learners exhibited native-like behaviours even though the gender marking is not encoded in their L1, while the advanced English-German learners did not. These findings indicated that late L2ers can eventually acquire unique-to-L2 knowledge and automatically use it in real time. And when reaching a near-native L2 proficiency level, late L2ers can use gender information of determiners as a cue for establishing predictive agreement relations in real-time comprehension and interpretation. This result argues against the Representation Deficit Account and supports the Performance Deficit Account and the Competition Model that late L2ers do not have deficient L2 representations: they can eventually acquire unique-to-L2 constructions and automatically use them during on-line predictive processing.

Along the same line, Hopp (2015) conducted a Visual World Paradigm study to investigate whether late L2ers can integrate case marking and verb semantics to generate anticipations during on-line processing. Forty-five English-German learners were recruited as L2 participants, 15 of them were low to intermediate proficiency L2ers, 19 of them were intermediate to high proficiency L2ers, and 11 of them were advanced L2ers. Twelve native German speakers participated as a control group. Case marking

can be treated as a unique-to-L2 construction to English-German learners since it only exists in German. Both nominative and accusative determiners were used in the materials. The example of their materials is illustrated in (2).

(2) a. Der Wolf tötet gleich den Hirsch.

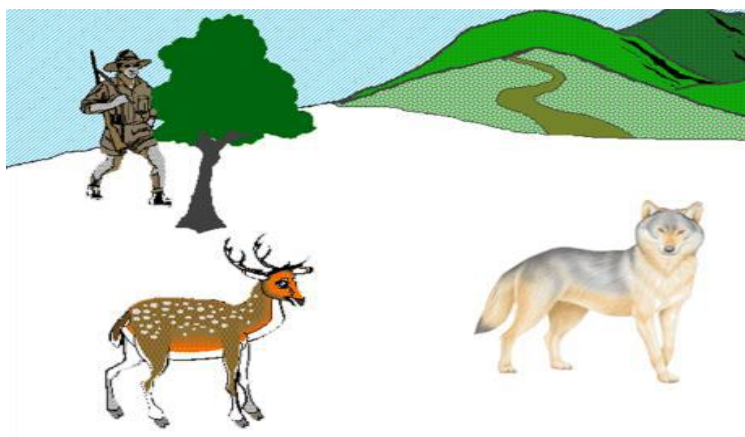
The<sub>NOM</sub> wolf kills soon the<sub>ACC</sub> deer

‘The wolf will soon kill the deer.’

b. Den Wolf tötet gleich der Jäger.

The<sub>ACC</sub> wolf kills soon the<sub>NOM</sub> hunter

‘The hunter will soon kill the wolf.’



Hopp (2015)

There are two areas of interest: the agent and the patient. By analysing how participants' fixations in these two interest areas change along with the input of the audio materials, the authors could tell how participants use case markings on the determiners to build anticipations during real-time processing. They found that native German speakers integrated case marking and verb semantic information in anticipatory processing, while late English-German learners relied more on verb semantic information for prediction, but not case marking information. Similar to what has been observed in Hopp (2012), in this study he found that the increase of the L2 proficiency affects L2ers' on-line use of morphosyntax information (case marking): compared to the low-intermediate group, the intermediate-high and the advanced group exhibited more native-liked behaviour patterns. Anticipation building and automatic use of unique-to-L2 knowledge can be

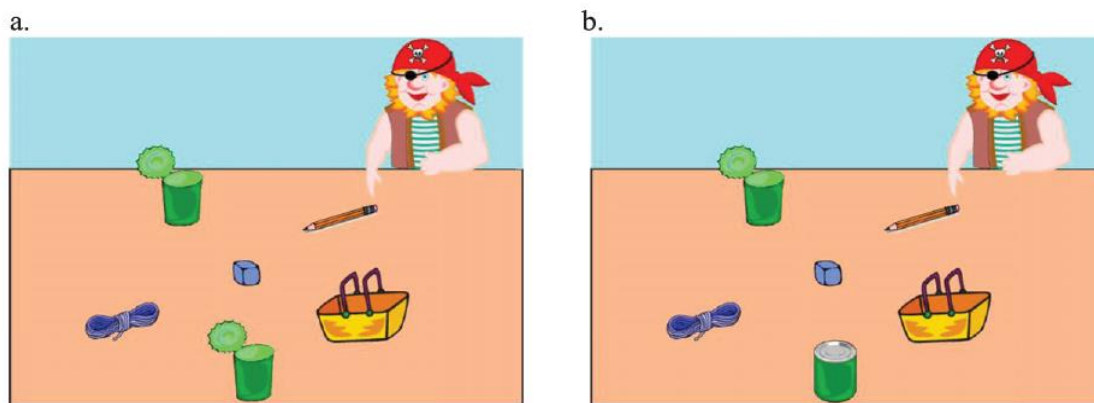
found in highly proficiency L2ers. These findings indicated that late L2ers can eventually acquire unique-to-L2 knowledge and integrate it during real time processing, which supports the Performance Deficit Account. Furthermore, both Hopp (2012) and Hopp (2015) found an L2 proficiency effect on late L2ers' predictive processing, which has also been observed by various studies (Chambers & Cooke, 2009; Dussias et al., 2013; Hopp, 2013; Leal, Slabakova, Farmer, 2017). These findings indicated that late L2ers' predictive processing difficulties are not caused by their 'general reduced predict ability', since if so, increased L2 proficiency would not affect their predictive performance. Rather, it is restricted by their proficiency. As long as they have reached an advanced L2 proficiency level, it is possible for late L2ers to build anticipations during on-line processing in a native-like manner.

The native-like behaviours in real-time anticipation-building have also been found in intermediate Mandarin-English learners. Trenkic, Mirkovic & Altmann (2013) used a Visual World Paradigm experiment to test Mandarin-English learners' on-line comprehension of English articles. Articles can be treated as a unique-to-L2 structure for Mandarin-English learners since they only exist in English but not Mandarin. Forty-eight intermediate Mandarin-English adult learners were recruited. Also, 56 native English speakers were recruited as a control group. Mandarin is an 'article-less' language compared to English, and as a consequence, Mandarin-English learners often omit articles or choose an inappropriate article for the context in previous studies (Ionin, Ko & Wexler, 2004; Jarvis, 2002; Luk & Shirai, 2009; Trenkic, 2002, 2007). In order to look into how late Mandarin-English learners comprehend English articles on-line in well-formed materials, and how they use article information to build anticipations, the authors manipulated the experimental materials to have two conditions: a two-compatible referent condition and a one-compatible referent condition. To be specific, the definiteness status of the target nominal phrase was manipulated in the audio materials ('The [agent] will put the [theme] inside **the/a** [goal]'), while the pragmatic affordances in the scene were manipulated in the visual materials. An example of the materials is illustrated in (3).

(3) Definite article: The pirate will put the cube inside **the** can.

Indefinite areicle:The pirate will put the cube inside **a** can.





*(In picture a there are two possible referents, e.g. the two opened cans; while in picture b there is only one possible referent, e.g. the opened can.)*

Trenkic, Mirkovic, & Altmann (2013)

Participants were required to watch some pictures on the screen and at the same time, listen to the descriptions about what is going to happen in the picture from headphones, and then to mouse-click on the location on the screen where the described object will end up when they finish listening. The authors expect that if late Mandarin-English learners can take advantage of articles in English as predicted by the Competition Model, they should start to land more fixations on the target faster in the audio-visual matched conditions (i.e. definite article *the* + one-compatible referent picture; indefinite article *a* + two-compatible referent picture) than in the audio-visual mismatched conditions (i.e. definite article *the* + two-compatible referent picture; indefinite article *a* + one-compatible referent picture). On the other hand, if late Mandarin-English learners cannot process articles on-line and rely overly on pragmatic cues, they should react faster when there is only one possible referent in the picture (e.g. picture b) than when two referents are available (e.g. picture a), regardless of the definiteness status of the target nominal phrase in the audio materials.

The results showed that native English speakers exhibited exactly the behaviour patterns expected. As for intermediate Mandarin-English learners, even though slower than native English speakers, they also exhibited similar patterns which indicated that they are able to make use of articles in English to build anticipations during on-line processing. In general, Trenkic et al. (2013) found that intermediate Mandarin-English learners can use English articles in real time to constrain referential domains and resolve reference efficiently. The results indicated that some unique-to-L2 morphosyntactic structures can be processed in a native-like manner in on-line processing and interpretation, as

predicted by the Competition Model. Also, these results offer some supporting evidence for the Performance Deficit Account that late L2ers still have full access to UG, and they can eventually acquire unique-to-L2 constructions.

In general, recent studies using the Visual World Paradigm found evidence that intermediate to advanced L2ers can use unique-to-L2 constructions to form predictions during on-line processing. Late L2ers do not have a ‘reduced predictive ability’. Their predictive performance improves along with increased L2 proficiency. Based on these observations and analyses, I argue that late L2ers can eventually use unique-to-L2 constructions to build anticipations during on-line processing. Increased L2 proficiency will facilitate their on-line predictive processing. In the current study, I predict that high proficiency Dutch-Mandarin learners are able to automatically use mass/count syntactic cues to build anticipations during on-line processing.

In the field of second language acquisition (SLA), in spite of the possible reasons for late L2ers’ processing difficulties and non-native-like behaviours discussed in Section 1 and Section 2, there is another possible factor which may lead to the difference between L2ers and native speakers: their different ways of learning languages (implicit learning vs. explicit learning). Native speakers acquire their L1 mostly through implicit learning, while most L2ers learn their L2 through explicit learning (in classrooms, with explicit and specific instructions, etc.). The way people learn languages (explicit vs. implicit) could have an effect on their language representations and performance (Krashen, 1981). In the next section, previous studies which explored implicit learning in L2 acquisition are introduced and discussed.

### **3. Implicit learning in L2 acquisition**

A potential factor which may affect late L2ers’ automatic use of L2 knowledge is their different ways of learning L2 from native speakers. In the field of SLA, there is a lot of interest in the analysis of implicit and explicit learning (Ellis, 1994; Hulstijn, 2005). Krashen (1981) proposed that there are two different ways of developing knowledge of L2: language acquisition (implicit learning) and language learning (explicit learning). Implicit learning is an incidental process which does not involve language learners’ attention and effort, just like how children learn their native languages; while explicit

learning is an intentional process, involving language learners' awareness and consciousness (Shanks & St. John, 1994; Williams, 2005, 2009).

The Critical Period Hypothesis (Lenneberg, 1967) claims that after puberty, language learners start to lose their ability to attain 'automatic acquisition from mere exposure' (Lenneberg, 1967:176). Late L2ers have to explicitly learn their L2, which is different from native children who implicitly acquire their L1. Compared to native children who acquire the language unconsciously, automatically and effortlessly, late L2ers learn L2 through a conscious and laboured effort. And as discussed in Section 1, due to this difference between L1 acquisition and L2 learning, the Fundamental Difference Approach which is based on the Critical Period Hypothesis claims that late L2ers have different L2 representations from native speakers, and cannot exhibit native-like performance.

However, previous studies have found some experimental results which cannot be explained by the Critical Period Hypothesis. Several studies observed implicit learning in artificial materials learning (Leung & Williams, 2014; Rebuschat & Williams, 2011) and second language acquisition (Cleary & Langle, 2007; Donaldson, 2011; Montrul & Slabakova, 2003; Rebuschat & Williams, 2011; Robinson, 2005; Tolentino & Tokowicz, 2014).

Rebuschat & Williams (2011) used a semi-artificial language, grammaticality judgments and subject measures of awareness to explore whether late L2ers can acquire L2 syntax implicitly. They formed an artificial language by following the grammatical word-order rules of German, and using meaningful English words as lexicons. Thirty-five native English speakers participated in the training and the post-test. In the training phase, the semi-artificial materials were displayed to participants through audio inputs. Participants were required to judge whether each sentence they heard is semantically plausible or not. In the post-testing phase, participants were asked to decide on each sentence's grammaticality and to report how confident they were in their judgment on a binary scale (low vs. high confidence). Their accuracy in the grammaticality judgment task was used as a measure of implicit learning. The results indicated that incidental exposure to L2 syntax can result in unconscious knowledge, implicit learning could happen in SLA.

In order to explore the potential role of L1 knowledge on implicit language learning, Leung & Williams (2014) conducted three experiments on native English speakers and

native Cantonese speakers using semi-artificial materials. In the training phase participants were exposed to visually presented noun phrases (the article and noun combinations). All the articles were the same between two groups of participants, while the nouns were presented in each participant group's L1 (English or Chinese). The explicit rule was a distance mapping between the article and the noun (far vs. near), while the hidden rule varied in three experiments: in Experiment 1 it was a mapping between the article and the animacy of the noun, in Experiment 2 it was a mapping between the article and an unnatural linguistic feature of the noun (the number of strokes in a Chinese character or the number of capital letters in an English word), in Experiment 3 it was a mapping between the article and the shape of the object the noun denoted (which is a feature derived from the classifier-noun combinations in Chinese). Participants were required to indicate a certain article's meaning (far or near) by pressing corresponding buttons. After the training phase, a questionnaire was used to probe participants' awareness of the hidden relations between the articles and the nouns. Participants' learning was measured by the increased reaction times when the hidden grammatical rule was violated compared to the grammatical version. The results showed that the implicit learning happened in both groups when the mapping between the article and the noun was semantically salient (the animacy of the nouns in Exp1). However, implicit learning was found only on native Cantonese speakers when the article-noun mapping was derived from the classifier-noun relations in Chinese (in Exp3). When the mapping between the article and the noun was an unusual linguistic concept (in Exp2), no evidence of implicit learning was spotted in either group. These findings indicated that salient semantic features are easier to be implicitly learned compared to unnatural linguistic features. And also language learners' L1 background affects their implicit learning: the structures which have similar instantiations in L1 can be implicitly learned.

Tolentino & Tokowicz (2014) investigated the effect of cross-language similarity on L2 learning with different methods of instruction by teaching English native speakers a subset of Swedish. The Swedish materials were manipulated to have three levels of cross-language similarity with English: similar, dissimilar, and unique-in-Swedish. Following Tokowicz & Whinney (2005), the authors considered a morphosyntactic feature as similar if it exists in both English and Swedish with similar experience (e.g. demonstrative determiner-noun number agreement), dissimilar if it exists in both English and Swedish but is instantiated differently (e.g. singular noun phrase definiteness

marking), and unique-to-L2 if it is absent in L1 (e.g. indefinite singular article-adjective gender agreement). Based on the Competition Model, they predicted that the similar materials should be associated with higher scores in the Grammatical Judgement Test (GJT), the dissimilar materials should be associated with lower GJT scores, and the unique-to-L2 features should have above-chance GJT performance since their uniqueness assigns them high cue strength and leaves no competition from L1. Also, the methods of instruction were divided into three groups: the Salience group in which contrast and colour highlighting were offered (explicit learning group), the Rule & Salience group in which both contrast and grammatical explanations were offered (explicit learning group), and the Control group which only had the material exposure (implicit learning group). The results showed that both cross-linguistic similarity and introduction methods affected participants' learning: the dissimilar materials had higher GJT scores in the Salience group and the Rule & Salience group than the Control group. The Rule & Salience group resulted in the highest score on the unique-to-L2 features. Further, there is an interesting finding that the Control group exhibited an above-chance performance in GJT, indicating that implicit learning happened in SLA. Within the Control group, the effect of cross-linguistic similarity was revealed: the similar features were associated with better performance than the dissimilar and unique-to-L2 features. To conclude, they found that cross-language similarity is associated with language learners' performance in GJT with similar features evoking better performance than dissimilar ones. Also, explicit instructions and grammatical rule explanations were especially useful in learning unique-to-L2 features. Moreover, compared to the dissimilar and the unique-to-L2 features, the similar features were easier to be implicitly learned. Even though it is difficult, unique-to-L2 features can be implicitly learned indicated by the above-chance scores in GJT.

Although these studies found that L1 has some effects on late L2ers' implicit learning, some other researchers found that unique-to-L2 constructions can also be implicitly learned by late L2ers. Donaldson (2011) found that near-native English-French learners demonstrated native-like behaviours in the production of left dislocation, which is a linguistic construction that cannot be found in English and is rarely explicitly taught in language classrooms. English-French learners can only pick it up by implicitly learning in an immersive French environment. Their native-like performance in spontaneous production indicated that even after puberty, late L2ers can still implicitly learn unique-

to-L2 constructions. Montrul & Slabakova (2003) investigated advanced English-Spanish learners' use of Preterit and Imperfect past tenses, which only exist in Spanish but not English. By using a Truth Value Judgment Task, they found that very advanced English-Spanish learners can eventually exhibit native-like behaviours in spite of the poverty of the stimulus. These findings are consistent with the argument in Rothman (2008) and Slabakova (2006) that late L2ers can implicitly acquire some linguistic constructions in L2 which cannot be transferred from L1.

Based on these studies, I argue that implicit learning can happen in late L2 acquisition, and it is possible for late L2ers to implicitly learn unique-to-L2 constructions. Also, from the above analyses we can see that previous studies exploring late L2ers' implicit learning often use artificial or semi-artificial materials, and some off-line tests and the training & post-test paradigms. This is because L2 acquisition is heavily influenced by L1-based processing strategies (N.C. Ellis & Sagarra, 2011; MacWhinney, 2008) since late L2ers start to learn their L2 with existing linguistic knowledge and habits from L1. In order to minimise the effect of prior knowledge, artificial grammars (Reber, 1967) or semi-artificial materials (Leung & Williams, 2012, 2014; N. C. Ellis, 2005; Williams & Lovatt, 2005) were chosen. Furthermore, even though for some L2ers, there is some unique-to-L2 knowledge which can be used to exclude L1's transfer effect, most of the linguistic features have been highlighted and taught explicitly and specifically in language learning classes (e.g. gender markings in Hopp (2012), case markings in Hopp (2015), articles in Trenkic et al. (2013)). It is difficult to use natural language materials and on-line processing paradigms to investigate L2ers' implicit learning of unique-to-L2 knowledge. Moreover, the training & post-test paradigm previous studies used restricts language learners' learning time and experience: participants can only have around 45mins to 1 hour to learn the hidden grammar in a very unnatural condition. It is important to explore late L2ers' implicit learning in a long period under an immersive L2 environment and under a natural linguistic input condition. Additionally, it is interesting to find out whether implicit learning in L2 acquisition could result in implicit knowledge which can be automatically used during on-line processing.

To fill these gaps, the current research explored late Dutch-Mandarin learners' on-line processing of nominal phrases with different mass/count syntactic cues. Different from Dutch, in which the mass/count distinction is instantiated through plural markings and articles, the mass/count distinction is encoded in the classifier system in Mandarin

(Cheng & Sybesma, 1998, 1999, among others). For Dutch-Mandarin learners, the mass/count distinction marked in the classifier system is a unique-to-L2 construction considering the absence of the classifier system in Dutch. Further, the mass/count syntactic cues are implicit to native Mandarin speakers as well as Dutch-Mandarin learners (native Mandarin speakers automatically use them in daily life without being able to spell out the specific rules). Dutch-Mandarin learners have only been taught about the obligatory appearance of a classifier between a numeral and a noun, and some high-frequency classifier-noun pairings, but not about the mass/count associations of different classifiers or about the mass/count syntactic structures. They can only implicitly learn this knowledge (i.e. the mass/count syntactic cues) through immersive experiences of a Mandarin-dominant environment since no textbooks of Mandarin learning ever include this knowledge. The mass/count syntactic cues in Mandarin nominal phrases offer us ideal natural language materials to investigate late language learners' implicit learning of a unique-to-L2 construction under a natural condition. Detailed information about the uniqueness and implicitness of the mass/count syntactic cues are presented in Chapter 2. Furthermore, the research methods used in the current research (the Visual World Paradigm and the Reading for Comprehension task) can offer us a fine-grained index about participants' on-line processing, which can then reflect whether implicit learning in L2 acquisition will result in implicit knowledge. Detailed information about the research methods employed in the current research is introduced in Chapter 3.

### **4. The current research**

The aim of the current research is to investigate high proficiency late Dutch-Mandarin learners' on-line processing of mass/count syntactic cues in Mandarin nominal phrases. To be specific, the research interests are whether high proficiency late Dutch-Mandarin learners can eventually exhibit native-like behaviours in the predictive processing of unique-to-L2 constructions; whether tasks with different workloads affect their performance; whether they can acquire unique-to-L2 constructions through implicit learning.

By analysing previous studies, I argue that late L2ers' processing difficulties are caused by their performance deficit (i.e., limited cognitive resources), instead of their 'deficient'

representations. Even though starting to learn L2 later than puberty, late L2ers can acquire native-like L2 representations and eventually exhibit native-like behaviours. Unique-to-L2 constructions can be implicitly learned by late L2ers.

Based on the Representation Deficit Account, late Dutch-Mandarin learners in the current research are expected to exhibit non-native-like behaviours during on-line processing, since they started to learn Mandarin after puberty, which leads to deficient Mandarin representations. Also, based on the Morphological Congruency Hypothesis, late Dutch-Mandarin learners are expected to exhibit non-native-like behaviours in making use of mass/count syntactic cues during real time processing. Since the mass/count syntactic cues are unique-to-L2 constructions to them, they cannot have full representations of these constructions due to their ‘fossilized’ brain system and the negative transfer from their L1. On the other hand, according to the Performance Deficit Account, high proficiency late Dutch-Mandarin learners can exhibit native-like behaviours during on-line processing. And it is more likely to observe their native-like behaviours in the Visual World Paradigm experiment than in the Reading for Comprehension experiment since the former one needs a comparatively lower workload than the latter one. In addition, according to the Feature Reassembly Hypothesis, the mass/count syntactic cues can be fully acquired by late Dutch-Mandarin learners through a ‘feature reassembly’ procedure. Thus, looking into high proficiency late Dutch-Mandarin learners’ on-line processing of mass/count syntactic cues in the Visual World Paradigm experiment and the Reading for Comprehension experiment, one could tell whether late L2ers can eventually exhibit native-like patterns, whether their processing difficulties are due to their representation deficit or performance deficit, and whether they can implicitly learn unique-to-L2 knowledge.

Detailed information about the two eye-tracking experiments (the Visual World Paradigm experiment and the Reading for Comprehension experiment) are presented in Chapter 4 and Chapter 5 respectively. A general discussion about the results from these two experiments and a conclusion are offered in Chapter 6.



## **II. Classifiers in Mandarin Chinese**

The main aim of the current research, as stated in Chapter 1, is to investigate high proficiency late Dutch-Mandarin learners' on-line processing of unique-to-L2 constructions, and their ability to use these constructions in anticipation-building in real-time processing. In order to tackle these questions, this study employs materials which meet the following two basic requirements: first of all, the materials are unique-in-Mandarin constructions for Dutch-Mandarin learners. In other words, these constructions only exist in Mandarin but not in Dutch. Secondly, these constructions can only be implicitly learned through an immersive experience in a Mandarin-dominant environment, and without any explicit and specific instructions from textbooks or language-teaching classes. This chapter presents the 'unique' and 'implicit' materials used in the current research.

In this chapter, the differences of nominal phrases between languages like English, Dutch, and Mandarin are presented in Section 1. The evidence of the uniqueness of the classifier system and the mass/count syntactic cues in Mandarin for Dutch-Mandarin learners is provided in Section 2. At the end of this chapter, the implicitness of the mass/count syntactic cues is validated in Section 3.

### **1. English/Dutch vs. Mandarin**

Nominal phrases are constructed in different ways between languages like English and Dutch, and languages like Mandarin. Based on nouns' function of being arguments or predicates, Chierchia (1998a, 1998b) proposed the Nominal Mapping Hypothesis, which divides languages into three groups: N [+arg, -pred] languages in which all bare nouns are arguments (e.g. Mandarin Chinese), N [-arg, +pred] languages in which all nouns are predicates (e.g. Italian), and N [+arg, +pred] languages in which nouns can be either

arguments or predicates (e.g. English, Dutch). In the current study, I focus on the differences between English/Dutch and Mandarin Chinese.

In English and Dutch, number is reflected in the morphology of articles and nouns, while in Mandarin neither articles nor plural morphemes exist. Instead, a classifier is obligatory when combining a noun with a numeral (e.g. Borer, 2005; Cheng & Sybesma, 1998, 1999, 2008, 2012; Doetjes, 1997).

In English, singularity is encoded through the indefinite article ‘a’ and plurality is marked morphologically on nouns. When expressing singularity, a definite/indefinite article or the numeral ‘one’ is used to combine with a noun, as illustrated in (4).

(4) a. a cat

b. the cat

c. one cat

Plural morphemes ‘-s/-es’ are used to mark plurality on nouns, as the examples in (5).

(5) a. \*three cat

b. three cats

c. \*two box

d. two boxes

In (5a) and (5b), when expressing plurality, the plural morpheme ‘-s’ has to be added to the noun ‘cat’. Similarly, in (5c) and (5d), the plural morpheme ‘-es’ needs to be added to the noun ‘box’ to realize plurality. It should be noted that the plural morphemes ‘-s’ and ‘-es’ are two different allomorphs of the same morpheme reflecting a phonological process. Bare nouns with count readings cannot occur with numerals greater than one directly, as the ungrammaticality of the phrases illustrated in (5a) and (5c).

On the basis of overt singular and plural markings, in languages like English and Dutch, the mass/count distinction can be easily noticed for nouns, articles and quantifiers. For example, only count nouns can be accompanied by an indefinite article but not mass nouns, as illustrated in (6).

(6) a. a cat

b. a box

c. \*a water

d. \*an oil

In (6a) and (6b), the count nouns ‘cat’ and ‘box’ can directly follow the indefinite article ‘a’, while the mass nouns ‘water’ and ‘oil’ cannot, as the examples in (12c) and (12d) are unacceptable in the reading ‘a unit of water’ and ‘a unit of oil’.

When expressing plurality, count nouns in English need to combine with plural morphemes and occur with numerals directly, as examples in (7a) and (7b). Mass nouns, however, need appropriate measure words to be inserted following numerals. And it is the measure words which undergo inflectional plural changes instead of the mass nouns themselves, as exemplified in (7c) and (7d).

(7) a. cat → five cats

b. box → three boxes

c. water → three bottles of water

d. oil → two drops of oil

Also, in English, the quantifiers which are used to modify count nouns and mass nouns are different. Count nouns occur with ‘many’, as in (8a), while mass nouns occur with ‘much’, as in (8b).

(8) a. many/\*much cats

b. much/\*many water

Similar to English, both singularity and plurality are overtly marked in Dutch. In Dutch, the indefinite article ‘*een*’ is used to mark singularity, while plural morphemes ‘*-en/-s*’ are needed to mark plurality, as the examples illustrated in (9).

(9) a. *een boek* – a book

*drie boeken* – three books

b. *een tafel* – a table

*drie tafels* – three tables

In (9a), the plural morpheme ‘*-en*’ has to be added to the noun ‘*boek*’ to realize plurality. Similarly, in (9b), the noun ‘*tafel*’ has to take the suffix ‘*-s*’ to realized plurality. And when expressing singularity, the indefinite article ‘*een*’ needs to be used.

In Dutch, when counting mass nouns, a measure word has to be inserted between the numeral and the noun. And again it is the measure word which undergoes the inflection<sup>5</sup>. For example, in (10a) and (10b), *kaas* ‘cheese’ is a mass noun. In order to count it, a measure word ‘*stuk*’ needs to be inserted between the numeral and the mass noun. And when expressing plurality, the measure word ‘*stuk*’ needs to be its plural form ‘*stukken*’, but not the mass noun ‘*kaas*’.

- (10). a. *een stuk kaas*  
           a   piece   cheese  
       b. *twee stukken kaas*  
           two   pieces   cheese

Different from English and Dutch, Mandarin Chinese has no number morphology to mark singularity or plurality of nouns<sup>6</sup>. Bare nouns in Mandarin Chinese have great flexibility in the contextual interpretation with respect to numbers and (in)definiteness. For example, the bare noun *mao* ‘cat’ in (11) has at least four possible interpretations, i.e. ‘a cat’, ‘cats’, ‘the cat’ and ‘the cats’, when given the appropriate contexts.

- (11) *wo kan jian mao le*  
       I   watch see cat LE  
       a. ‘I saw a cat.’  
       b. ‘I saw cats.’  
       c. ‘I saw the cat.’  
       d. ‘I saw the cats.’

In nominal phrases, nouns never change their forms. The nominal phrases *yi zhi mao* ‘one CL cat’ in (12a) and *wu zhi mao* ‘five CL cat’ in (12b) only differ in their numerals but not the noun’s form. Even when there is a plural quantifier, such as *xuduo* ‘many’ in

---

<sup>5</sup> It should be noted that not all measure words in Dutch need to be pluralized when occurring with cardinal numbers. For example, some measure words like *kilo* ‘kilo’, *litter* ‘liter’, *gram* ‘gram’ do not overtly mark the singular/plural distinction. See Doetjes (1997): 189.

<sup>6</sup> Some researchers claim that ‘-men’ is a plural morpheme in Mandarin Chinese which can only occur with pronouns and nouns with [+human] feature (Choi, Ionin, Zhu, 2017; Ijic, 1994). In the current study, ‘-men’ is not under my research interest. I will let the question open that whether there is a plural morpheme in Mandarin Chinese.

(12c), the noun *mao* is used in its bare form. So compared to languages like English and Dutch, Mandarin Chinese is regarded to be a ‘number-neutral’ language (Rullmann & You, 2006).

(12) a. *yi zhi mao*

one CL cat

‘one cat’

b. *wu zhi mao*

five CL cat

‘five cats’

c. *xuduo mao*

many cat

‘many cats’

Another difference between nominal phrases in Mandarin Chinese and English/Dutch is that Chinese has neither definite nor indefinite articles in the nominal domain. Native speakers use demonstratives, such as *zhe* ‘this’ or *na* ‘that’ in Mandarin to express definiteness overtly, as showed in (13a) and (13b). Numeral phrases such as ‘numeral+ CL+ N’, are used to express indefiniteness as in (13c) and (13d).

(13) a. *wo hen xihuan ni song wo de na ge liwu*

I very like you give me DE that CL gift

‘I like the gift you gave me.’

b. *zhe ben shu shi Lili xie de*

this CL book be Lily write DE

‘This book is written by Lily.’

c. *Lily mai le yi liang xin che*

Lily buy LE one CL new car

‘Lily bought a new car.’

d. *yuanzi li you liang ke shu*

yard inside exist two CL tree

‘There are two trees in the yard.’

Generally speaking, Mandarin Chinese is a number-neutral and an article-less language. Bare nouns in Mandarin are open for either singular or plural, and either definite or indefinite interpretations. Thus unlike English or Dutch, the mass/count distinction cannot be overtly marked by number morphemes in Mandarin.

Mandarin Chinese is also regarded as a typical ‘classifier language’ because of the individual classifiers, which are not found in languages like English and Dutch. In Mandarin, nouns cannot combine directly with numerals: it is obligatory to insert a classifier between a numeral and a noun when a noun is counted, as exemplified in (14).

(14) a. \**san mao*

three cat

Intended reading: ‘three cats’

b. *san zhi mao*

three CL cat

‘three cats’

c. \**san shui*

three water

Intended reading: ‘three units (e.g. cups/bottles/bowls) of water’

d. *san bei shui*

three CL-cup water

‘three cups of water’

In (14), no matter whether the noun is a count noun (e.g. *mao* ‘cat’ in (14a) and (14b)) or a mass noun<sup>7</sup> (e.g. *shui* ‘water’ in (14c) and (14d)), it cannot be used with a numeral

---

<sup>7</sup> In the current research, the **mass/count noun** refers to the ontological concept of a noun, and not to a noun's grammatical status. In other words, it refers to the conceptual status of a noun denoted under standard circumstances. It should be noted that the mapping from concept to grammar with regard to the mass/count distinction is not always straightforward if we consider nouns like ‘furniture’, ‘jewelry’, etc. In the current study, these flexible nouns are excluded from materials. We only focus on the nouns which are normed and rated as typical count/mass nouns. Also, the debate about whether the

directly, as exemplified in (14a) and (14c). A classifier is obligatory whenever a noun is counted by a numeral, as in (14b) and (14d).

Classifiers in Mandarin Chinese are different from measure words in English/Dutch. For the most part, measure words in English/Dutch can only be used to describe the units of mass nouns, like ‘a piece of paper’, ‘two bottles of water’, or groups of count nouns with plural readings, like ‘a bunch of flowers’, ‘three boxes of apples’. And when the numeral is bigger than one, the measure words should be in plural forms. On the other hand, in Mandarin Chinese, a classifier is used with a number regardless whether the noun is a count noun with a singular/plural/portion reading, or a mass noun. And there is no morphological plural form for classifiers.

Even though the presence of a classifier is obligatory in every numeral phrase, it should be noted that the pairing of nouns and classifiers is not free in Mandarin Chinese, as illustrated in (15).

- (15) a. *yi ge/ke/\*tan/\*di pingguo*  
           one CL                   apple  
           ‘an apple’
- b. *yi di/tan/\*ge/\*ke shui*  
           one CL                   water  
           ‘a drop/pool/unit of water’
- c. *yi ge beizi/ bao/ \*you/ \*shazi*  
           one CL   cup/   bag/   \*oil/   \*sand  
           ‘a cup/bag/\*oil/\*sand’
- d. *yi di guozhi/ jiangyou/ \*pingguo/ \*gou*  
           one CL\_drop juice/ soy sauce/ \*apple/   \*dog  
           ‘a drop of juice/soy sauce/\*apple/\*dog’

In (15a), *pingguo* ‘apple’ can occur with the classifiers *ge* (an individual unit) and *ke* (an individual unit which usually occurs with small and round objects), but cannot occur

---

mass/count distinction is encoded at the lexical entry of nouns (Doetjes, 2007) or at the syntactic structure of nominal phrases (Borer, 2005) is not the concern of the current research.

with the classifiers *tan* ‘pool’ and *di* ‘drop’. While in (15b), *shui* ‘water’ can occur with the classifiers *tan* ‘pool’ and *di* ‘drop’, but not with the classifiers *ge* and *ke*. In (15c), the classifier *ge* can occur with nouns like *beizi* ‘cup’, *bao* ‘bag’, *pingguo* ‘apple’, but cannot occur with nouns like *you* ‘oil’, *shazi* ‘sand’, and *shui* ‘water’. In (15d), the classifier *di* ‘drop’ can occur with nouns like *guozhi* ‘juice’, *jiangyou* ‘soy sauce’, *shui* ‘water’, but cannot occur with nouns like *pingguo* ‘apple’ and *gou* ‘dog’.

Apart from the obligatory presence of classifiers and the restrictions of the pairing of classifiers and nouns, classifiers in Mandarin also differ from number morphology in Dutch and English in many other aspects.

The first difference can be observed when numerals are removed from the picture. When numerals do not occur, number morphology can still mark count nouns, and bare plural nouns can freely occur in argument positions in languages like English and Dutch, as illustrated in (16a) and (16b). However, when numerals are absent, the individual classifier marked noun – the ‘CL-noun’ combination – is only acceptable in postverbal positions in Mandarin, as shown in (16c) and (16d).

(16) a. Dogs are men’s best friends.

b. Lily made cookies by herself.

c. \**zhi gou yao guo malu*

CL dog want cross road

‘A dog wanted to cross the road.’

d. *Lily yang le zhi gou*

Lily keep LE CL dog

‘Lily kept a dog.’

Another difference between number morphology and classifiers emerges when both of them are absent along with numerals from the nominal phrases: bare nouns in these two kinds of languages behave differently. In English, bare count nouns without number morphology and without numerals cannot occur as bare arguments, as shown in (17a). In Mandarin, bare count nouns without numerals and without classifiers are grammatical expressions with different possible interpretations, as in (17b). And they can always



merge directly with a verb, and occur in either preverbal or postverbal positions, as seen in (17c) and (17d).

(17) a. \*Student, farmer and worker can receive a discount.

b. *xuesheng*

Student

‘a student/students/the students’

c. *xuesheng, nongmin he gongren keyi xiangshou youhui*

Student, farmer and worker can enjoy discount

‘Students, farmers and workers can have a discount.’

d. *wo kan jian xuesheng le*

I watch see student LE

‘I’ve seen a student/some students/the student/the students.’

Furthermore, number morphology is, to some extent, number specific, i.e. it is either singular or plural, as seen in (18a). However, classifiers are not, i.e. they remain the same regardless of singularity or plurality, as seen in (18b).

(18) a. one student/ two students

b. *yi ge xuesheng/ liang ge xuesheng*

one CL student/ two CL student

The last difference between number morphology and classifiers is that the classifier system is rich (e.g. there are different types of classifiers) and word-specific, which is different from the general plural morpheme *-s/-es*, as shown in (19).

(19) a. *san ge xuesheng*

three CL student

‘three students’

b. *wu zhi laohu*

five CL tiger

‘five tigers’

c. *liang ke shu*

two CL tree

‘two trees’

d. *wu duo hua*

five CL flower

‘five flowers’

From the examples in (19) we can see that, in Mandarin, different classifiers select different groups of nouns: animate creators use different classifiers from inanimate plants, while even within the animate/inanimate category, objects with different features (e.g. shape, size) are occurring with different classifiers. In (19a) and (19b), *xuesheng* ‘student’ and *laohu* ‘tiger’ are both animate creators, *xuesheng* occurs with *ge* which is a general classifier that can fit with a lot of nouns, while *laohu* occurs with *zhi* which is a typical classifier usually occurs with animals. In (19c) and (19d), *shu* ‘tree’ and *hua* ‘flower’ both are inanimate plants, they occur with different classifiers: *shu* occurs with *ke* which usually describes individual objects with a thin and long stick-like shape, while *hua* occurs with *duo* which is often specifically used with flowers.

Based on above observations, the number marking differences between English/Dutch and Mandarin are summarized in Table 2. The differences between number morphemes and the classifier system are summarized in Table 3.

Table 2. The number marking differences among English, Dutch, and Mandarin

	English	Dutch	Mandarin
<b>Singularity</b>	a	een	<ul style="list-style-type: none"> <li>• No article/plural form</li> <li>• Bare N</li> </ul>
	Count N: ‘-s/-es’	Count N: ‘-en/-s’	<ul style="list-style-type: none"> <li>• Classifiers are obligatory</li> </ul>
<b>Plurality</b>	Mass N: measure words	Mass N: measure words	<ul style="list-style-type: none"> <li>• Num+CI+N</li> <li>• Demonstrative+CI+N</li> </ul>

Table 3. The differences between number morphemes and the classifier system

	Number morphemes	Classifiers
<b>Number specific</b>	Pluralized when necessary	No plural form
<b>Relationship with nouns</b>	Measure words are obligatory for mass nouns	<ul style="list-style-type: none"> <li>• Obligatory appearance</li> <li>• Restrictions of Cl-N pairings</li> </ul>
<b>Without numerals</b>	√	×
<b>Bare nouns</b>	×	√
<b>Diversity</b>	Limited and general	Rich and word-specific

From Table 2 and Table 3 we can see that, Dutch and Mandarin are different from each other in the way of number markings in nominal phrases. The classifier system in Mandarin is very different from number morphemes in Dutch.

Considering the lack of overt number morphology in Mandarin, the mass/count distinction must be encoded differently from languages like English and Dutch. In the next section, the question of how the mass/count distinction is grammaticalized in Mandarin is discussed.

## 2. Mass/Count in Mandarin Chinese

The observations that bare nouns in Mandarin could have multiple flexible meanings, no plural morphemes are needed when expressing plurality, and a classifier is obligatory when counting nouns, lead some researchers to argue that nouns in languages like Mandarin Chinese are fundamentally different from nouns in languages like English and Dutch. These researchers further argue that the differences cross languages lead to people's different mass/count mental representations. By this view, the objects/substances in the world are situated along an 'individuation continuum', and different languages use different criteria to divide count objects from mass substances (Allan, Gentner & Boroditsky, 2001; Lucy, 1992).

Lucy (1992) found that when judging whether a noun is a count noun or a mass noun, English speakers prefer to use the shape as a salient dimension, while Yucatec Mayan speakers focus more on the material. Lucy argues that this difference in mass/count criteria is caused by speakers' different languages: 'use of the English lexical items

routinely draws attention to the shape of a referent insofar as its form is the basis for incorporating it under some lexical label. Use of Yucatec lexical items, by contrast, routinely draws attention to the material composition of a referent insofar as its substance is the basis for incorporating it under some lexical label' (Lucy 1992, P89). By using a word extension task, Imai & Gentner (1997) found that 2-years-old English speaking children were more likely to extend words on the basis of shape than 2-years-old Japanese speaking children. Based on the results, they argue that 'Japanese speakers and English speakers appeared to use different criteria in determining the class membership for a given instance, suggesting that they have a different representation for, or at least a different boundary between individuals and non-individuals' (Imai & Gentner 1997, P195).

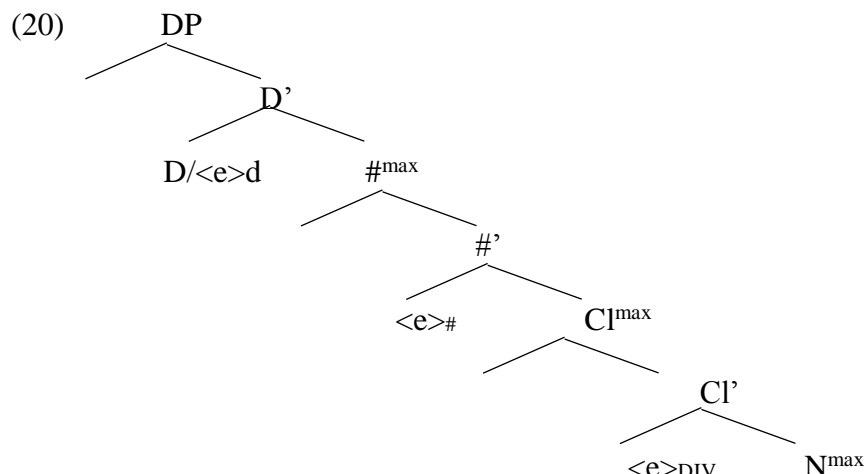
However, recently more and more research found evidence challenging this kind of neo-Whorfian Hypothesis. Mazuka & Friedman (2000) found that the difference between English- and Japanese-speaking people in judging count/mass nouns may be affected by different culture and education. Li & Gleitman (2002) and Gleitman & Papafragou (2005) argue that the observed cross-linguistic difference in word extension can also be explained by the difference of count/mass nouns' frequency. Li, Dunham & Carey (2009) found that the cross-linguistic difference in word extension disappeared when participants were asked to rate novel stimuli as objects or substances along a seven-point scale. They interpreted that this finding indicated that even though the mass/count distinction is grammaticalized in different ways in varied languages, there is no fundamental difference in how objects are construed. Speakers of different languages share the same mental representations of count/mass entities. Using a word rating task and a quantity judgment task on native English speakers and native Japanese speakers, and a word extension task on Mandarin-English learners, Barner, Inagaki & Li (2009) found that native speakers of English and Japanese, and bilingual speakers of Mandarin and English all access a universal ontology of individuals no matter what language they were tested on. They claim that even though different languages mark the mass/count distinction in different ways, people share the same count/mass mental representations. Cross-linguistically, some nouns denote things that have individual and solid units, while some nouns denote things that can only be construed as unindividuated stuff. 'In between these two poles are a host of words that provide multiple construals, each word differing in the extent to which one construal is favoured over the other. Languages, we assert, do

not alter preferences between construals, but rather differ in whether they provide obligatory syntactic mechanisms for selecting between them' (Barner et al 2009, P11). For example, languages like English and Dutch provide overt mass/count syntax to select appropriate meanings from the construal (i.e. the plural morphemes, articles, quantifiers). While in languages like Mandarin, there is no obligatory grammatical distinction on nouns. Languages only differ in their method/capacity to tease apart lexical construals. Choi, Ionin & Zhu (2017) further support the view that different languages do not shape people's perception and conception differently. Conducting a grammar task on native speakers of Korean-English learners and Mandarin-English learners, they found that the cross-linguistic differences among these three languages did not affect speakers' performance on count/mass judgment. Both Korean-English learners and Mandarin-English learners exhibited similar patterns of the count/mass judgments. They share the same conceptual/semantic representations of count/mass entities.

Based on these previous studies, in the current study, I argue that even though nouns in Mandarin behave differently from nouns in English/Dutch, speakers of these languages share the same count/mass criteria and perceptions of objects/substances in the world.

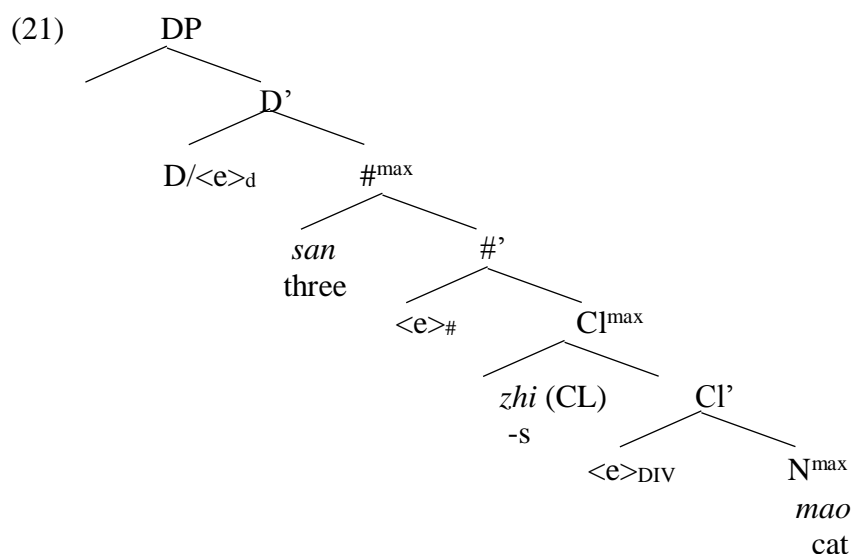
Since the mass/count distinction is already built in people's mental representations, and they share the same perceptions of entities in the world, then what needs to be acquired by Dutch-Mandarin learners when they learn the mass/count distinction in Mandarin grammar? What is the difference of the grammaticalized mass/count distinction between Dutch and Mandarin? In the current study, based on Borer (2005), I argue that the formal syntactic structures of count/mass phrases are shared across different languages. What Dutch-Mandarin learners need to acquire is not a new functional category, instead, they need to learn a new feature value. To be specific, they need to acquire a new way of expressing the semantic count/mass distinction.

Borer (2005) proposed that the mass/count distinction is grammatically marked in both numeral languages (e.g. English and Dutch) and classifier languages (e.g. Mandarin). She assumed that the internal structure of a nominal phrase has three functional layers above the bare noun: the classifier phrase ( $CI^{max}$ ), the quantity phrase ( $\#^{max}$ ) and the determiner phrases (DP). And also, nominal phrases in all languages share this uniform structure, as shown in (20).



The classifier head possesses the open value  $\langle e \rangle_{\text{DIV}}$ , with  $\text{DIV}$  standing for the ‘dividing’ function. Overt classifiers, plural inflection, and indefinite articles are assumed to base-generate in the  $\text{CI}^{\text{max}}$  domain and assign range to  $\langle e \rangle_{\text{DIV}}$  to accomplish the portioning-out function. The difference between the overt classifiers and plural markings is ‘the plural marker is a spell-out of an abstract head feature  $\langle \text{div} \rangle$  on a moved N-stem, while the classifier is an independent f-morph’ (Borer, 2005: 95).

Specifically, in numeral languages with overt determiners, such as English, both plural inflections and the indefinite determiner can assign a range to the open value  $\langle e \rangle_{\text{DIV}}$  and accomplish the portioning out function. In classifier languages like Mandarin, it is the classifiers that assign a range and accomplish that function. Thus according to Borer, nominal phrases in languages like Chinese and languages like English/Dutch share the same structure as illustrated in (21).



(Borer, 2005: 96)

A key element in Borer's nominal structure is that the plural inflection equals to the classifier system. Specifically, the plural inflection is assumed to be a distinct instantiation of the classifier system and appear in the same position as classifiers.

Similarly, Doetjes (1997) argued that classifiers in languages like Chinese share the same role with plural morphemes in languages like English, both of them can indicate the presence of proper countable units. To be specific, in order to be counted directly, the partitioning or unit of what nouns denote must be syntactically visible, and such syntactically visibility is marked in the classifier system in Mandarin, but via plural morphemes and the indefinite article in English. In this case, classifier system in Chinese has the parallel function with plural morphemes in English. In a similar vein, Cheng & Sybesma (1998, 1999) claim that the mass/count distinction is encoded in the classifier system in Mandarin.

According to Cheng & Sybesma (1998, 1999), classifiers in Mandarin can be roughly divided into two groups based on their functions: count-classifiers which simply name the unit that entities naturally have; and massifiers which create units to measure substances and pluralities. There are some distributional differences between count-classifiers and massifiers: only massifiers can be followed by the modification marker *de*, and only massifiers can be modified by adjectives like *da* 'big' and *xiao* 'small'.

Following Tang (1990), Cheng & Sybesma (1998) argue that the modification marker *de* can appear optionally between a massifier and a noun with a mass reading or a plural reading, but cannot occur between a count-classifier and a noun (also see Chao, 1968; Paris, 1981). Examples are shown in (22).

- (22) a. *san bei (de) shui*  
           three CL-cup DE water  
           'three cups of water'
- b. *san xiang (de) shu*  
           three CL-box DE book  
           'three boxes of books'
- c. *liang tou (\*de) niu*

two CL-head DE cow

‘two cows’

d. *wu jian (\*de) chenshan*

five CL DE shirt

‘five shirts’

*Bei* ‘cup’ in (22a) and *xiang* ‘box’ in (22b) are massifiers since they represent measurement units and describe the containers of the entities. The modification marker *de* can appear optionally following these massifiers. However, it is ungrammatical to insert *de* after the classifiers in (22c) and (22d), since *tou* ‘head’ in (22c) and *jian* ‘an individual classifier which usually occurs with clothing’ in (22d) are count-classifiers since they just name the units naturally possessed by the objects denoted by nouns.

Only massifiers can be modified by adjectives like *xiao* ‘small’ and *da* ‘big’, not count-classifiers, as illustrated in (23).

(23) a. *san da zhang zhi*

three big CL-piece paper

‘three big pieces of paper’

b. *san xiao bei shui*

three small CL-cup water

‘three small cups of water’

c. \**san da zhi gou*

three big CL dog

Intended reading: ‘three big dogs’

d. *san zhi da gou*

three CL big dog

‘three big dogs’

e. \**si xiao ge xiangzi*

four small CL box



Intended reading: ‘four small boxes’

f. *si ge xiao xiangzi*

four CL small box

‘four small boxes’

In (23a) and (23b), the adjective *da* ‘big’ and *xiao* ‘small’ can modify massifiers *zhang* ‘piece’ and *bei* ‘cup’. However, if adjectives are added before count-classifiers like *zhi* in (23c) and *ge* in (23e), the nominal phrases become ungrammatical. When the classifiers are count-classifiers, adjectives can only modify the nouns and appear between count-classifiers and nouns, as seen in (23d) and (23f).

It should be noted that the insertion of *de* and the ability to be modified by adjectives are determined by the types of classifiers (i.e. either count-classifier or massifier), but not by the mass/count status of nouns. Examples are shown in (24).

(24) a. \**san ge de ren*

three CL DE person

‘three people’

b. \**san da ge ren*

three big CL person

‘three big people’

c. *san qun de ren*

three CL-crowd DE person

‘three crowds of people’

d. *san da qun ren*

three big CL-crowd person

‘three big crowds of people’

In (24), the noun *ren* ‘person’ remains the same in all four cases, but the classifiers are different. In (24a) and (24b), the count-classifier *ge* cannot be followed by the modification marker *de*, neither can it be modified by the adjective *da* ‘big’. While in

(24c) and (24d), the massifier *qun* ‘crowd’ can be followed by the modification marker *de* and can be modified by *da* ‘big’.

Count-classifiers also differ from massifiers on the restrictions of the classifier-noun pairing. The count-classifiers/massifiers distinction does not mark, strictly, the mass/count distinction. Count-classifiers usually occur only with count nouns with countable singular readings, while not all massifiers are exclusively used with mass nouns. Some massifiers can occur with either mass nouns or count nouns with plural/portion readings. Examples are illustrated in (25).

(25) a. *yi ge pingguo/ \*shui*

one CL apple/ \*water

‘an apple/\*a water’

b. *yi zhi mao /\*you*

one CL cat/ \*oil

‘a cat/\*a oil’

c. *yi ping shui/ ganlan*

one CL\_bottle water/ olive

‘a bottle of water/olives/\*olive’

d. *yi dui shazi/ pingguo*

one CL\_pile sand/ apple

‘a pile of sand/apples/\*apple’

e. *yi kuai nailao/ pingguo*

one CL\_chunk cheese/ apple

‘a chunk of cheese/apple’

f. *yi pian mianbao/ huanggua*

one CL\_slice bread/ cucumber

‘a slice of bread/cucumber’

In (25a), *ge* is a typical count-classifier, and it can only occur with nouns which have default units, such as *pingguo* ‘apple’ in this case. It cannot occur with nouns which do

not have specific units like *shui* ‘water’. Similarly, in (25b), the count-classifier *zhi* can only occur with *mao* ‘cat’ which has discrete units, but not *you* ‘oil’ which does not. In (25c), *ping* ‘bottle’ is a massifier and can occur with *shui* ‘water’ which does not have specific units. It can also occur with *ganlan* ‘olive’, which semantically is a count noun and expresses a plural reading in this specific case. Likewise, in (25d), the massifier *dui* ‘pile’ can occur with mass nouns like *shazi* ‘sand’, and it can also occur with count nouns like *pingguo* ‘apple’ which has a plural reading in this case. In (25e), the massifier *kuai* ‘chunk’ can occur with mass nouns like *naillao* ‘cheese’, and can also occur with count nouns like *pingguo* ‘apple’, which has a portion meaning (i.e. ‘a chunk of apple’). In (25f), the massifier *pian* ‘slice’ can occur with mass nouns like *mianbao* ‘bread’, as well as count nouns like *huanggua* ‘cucumber’ which has a portion meaning (i.e. ‘a piece of cucumber’).

It should be noted that even though count nouns can occur with either count-classifiers or massifiers, they have different meanings when occurring with different classifiers: the count noun *pingguo* ‘apple’ expresses a countable meaning (‘individual apple’) when occurring with a count-classifier *ge* in (25a), a plural meaning (‘apples’) when occurring with a massifier like *dui* ‘pile’ in (25d), and expresses a portion reading (‘a part of apple’) when occurring with a massifier like *kuai* in (25e).

Additionally, Cheng & Sybesma (1998) argue that even though the modification marker *de* can optionally occur after massifiers, nominal phrases have different interpretations with and without *de*. The insertion of *de* after a massifier subtly changes the interpretation of a nominal phrase to emphasize the total volume of the substance. When nominal phrases have the sequence [Num+Cl+N] with a massifier, they express mass or plural meanings in which the measure unit the massifier denoted is concretely present. When nominal phrases have the sequence [Num+Cl+*de*+N] with a massifier, they express quantity meanings in which the measure unit the massifier denoted could be abstract. Examples are shown in (26).

(26) a. *Lily he le san ping jiu*

Lily drink LE three CL\_bottle wine

‘Lily drank three bottles of wine’

b. *Lily he le san ping de jiu*

Lily drink LE three CL\_bottle DE wine

‘The volume of the wine Lily drank is three bottles’

In (26a), when the nominal phrase *san ping jiu* ‘three bottles of wine’ has the sequence [Num+Cl+N], the container the massifier *ping* ‘bottle’ denoted must be concretely present. In (26b), however, when the nominal phrase *san ping de jiu* has the sequence [Num+Cl+*de*+N], the container ‘bottle’ could be abstract, and the whole nominal phrase conveys a meaning of ‘three bottles worth of wine’, instead of ‘three real bottles of wine’. Thus, massifiers have slightly different meanings and properties with and without the modification marker *de*.

To sum up, even though there is no number marking on nouns, nor definite/indefinite articles in Mandarin Chinese, the mass/count distinction can be reflected at the classifier level by dividing classifiers into two groups: count-classifiers and massifiers. Count-classifiers usually occur with count nouns with individual readings, while massifiers occur with either count nouns with plural/portion readings or mass nouns. In addition, the ability to be modified by adjectives and the insertion of the modification marker *de* indicate a massifier reading (Cheng & Sybesma, 1998; Li & Rothstein, 2012). Count nouns with different classifiers express different meanings, and massifiers with different syntactic cues (e.g. with and without *de*) have different properties. The differences between count-classifiers and massifiers are illustrated in Table 4.

Table 4. The differences between count-classifiers and massifiers

	Count-Classifier	Massifier
<b>Semantic function</b>	Naming units objects naturally have	Creating ways of measurement
<b>Insertion of <i>de</i></b>	×	√
<b>Ability to be modified</b>	×	√
<b>Cl-N pairing</b>	Count nouns only	Both count and mass nouns

The markings of count singular, count plural, and mass in English/Dutch and Mandarin are summarized in Table 5.

Table 5. Different mass/count markings in English/Dutch and Mandarin

English/Dutch	Mandarin
---------------	----------

<b>Count</b>	a + count bare noun	• count bare noun
<b>Singular</b>		• yi + count-classifier + count noun
<b>Count Plural</b>	(numeral) + count noun + plural morphemes	• count bare noun • numeral + count-classifier + count noun • numeral + massifier + count noun
<b>Mass</b>	• mass bare noun • (numeral) + measure word + plural morphemes + mass noun	• mass bare noun • numeral + massifier + mass noun

It should be noted that, apart from count-classifiers and massifiers, in Mandarin Chinese, there exist some ‘dual-role’ classifiers: they can be count-classifiers when naming the individual unit of the objects denoted by count nouns, but they can also be massifiers which create a way of measuring substances or materials, as shown in (25).

- (27) a. *yi ba jiandao/ yaoshi/ shaozi/ shazi/ shizi*  
 one CL\_handful scissors/ key/ spoon/ sand/ pebble  
 ‘one pair of scissors, a key/spoon, a handful of sand/pebbles’
- b. *yi kuai shoujuan/ yupei/ feizao/ dangao*  
 one CL\_chunk handkerchief/ shaped jade/ soap/ cake  
 ‘a handkerchief/shaped jade, a bar of soap, a chunk of cake’
- c. *yi gen huanggua/ xiangjiao/ toufa/ mugun*  
 one CL\_rod cucumber/ banana/ hair/ stick  
 ‘a cucumber/banana, a (string of) hair, a stick’
- d. *yi pian guangpan/ shuye/ mianbao/ nailao*  
 one CL\_slice disk/ leaf/ bread/ cheese  
 ‘a record/leaf, a piece of bread/cheese’

In (27a), the classifier *ba* ‘handful’ can occur with both count nouns like *jiandao* ‘scissors’, *yaoshi* ‘key’ and mass nouns like *shazi* ‘sand’. When it occurs with a count noun such as *yaoshi* ‘key’, it describes the unit a key naturally has. When it occurs with a mass noun such as *shazi* ‘sand’, it creates a way (e.g. handful) to measure the amount of the sand. In (27b), the classifier *kuai* ‘chunk’ describes the individual unit of a handkerchief or a shaped jade when it occurs with the count nouns *shoujuan* ‘handkerchief’ and *yupei* ‘shaped jade’. When it occurs with mass nouns like *feizao* ‘soap’ and *dangao* ‘cake’, the massifier *kuai* indicates that ‘soap’ and ‘cake’ should be measured by chunks. In (27c), the classifier *gen* ‘rod’ can occur with count nouns like *huanggua* ‘cucumber’ and *xiangjiao* ‘banana’, as well as mass nouns like *toufa* ‘hair’. *Gen* describes the unit a cucumber or a banana naturally has, but also creates a unit ‘rod’ to measure ‘hair’. In (27d), the classifier *pian* ‘piece’ can occur with count nouns like *guangpan* ‘disk’ and *shuye* ‘leaf’, mass nouns like *mianbao* ‘bread’ and *nailao* ‘cheese’. When occurring with *guangpan* ‘disk’ and *shuye* ‘leaf’, *pian* simply names the default unit of a disk and a leaf, but when it occurs with ‘bread’ and ‘cheese’, it indicates that the counting unit should be ‘slice’.

Since they possess both count-classifier meanings and massifier meanings, these dual-role classifiers can be treated as ambiguous words. Different structures of nominal phrases (classifiers being modified by adjectives or not, with or without the insertion of *de*), based on Cheng & Sybesma (1998, 1999), can be treated as syntactic cues for either count-classifier or massifier meanings. Thus, using these dual-role classifiers in the nominal phrase with different mass/count syntactic cues, I investigate how native Mandarin speakers and high proficiency late Dutch-Mandarin learners take advantage of these syntactic cues during real-time sentence processing.

In order to accomplish this task, Dutch-Mandarin learners need to acquire both the lexical knowledge and the syntactic knowledge of the classifier system in Mandarin. In the current study, the lexical knowledge of classifiers including which type of nouns a classifier prefers to occur with (as discussed in (25)), and what possible meanings a classifier could have (as discussed in (27)). The syntactic knowledge of classifiers including the possibility of being modified by adjectives (as discussed in (23)), and the capacity of being inserted the modification marker *de* (as discussed in (24)).

Based on the discussions in the beginning of this section, speakers of Dutch and Mandarin share the same count/mass mental representations, which means they do not

need to learn a new semantic mass/count distinction. Also, along with Borer (2005) I argue that the formal syntactic structure of nominal phrases are the same cross different languages, indicating that Dutch-Mandarin learners do not need to learn a new functional category. Based on Cheng & Sybesma (1998, 1999) and Doetjes (1997), the classifier system in Mandarin and the plural morphemes and the indefinite articles in English/Dutch have the parallel function and could be treated as different expressions of the same feature. Thus, what Dutch-Mandarin learners are facing when acquire the mass/count distinction in Mandarin is learning a new way to express the mass/count distinction. They need to acquire both the lexical knowledge and the syntactic knowledge of classifiers in Mandarin. And both the lexical and the syntactic knowledge of the classifier system is unique-to-Mandarin since there is no classifier in Dutch.

To be specific, in order for Dutch-Mandarin learners to accomplish the task, firstly they need to know that the mass/count distinction is encoded in different ways between Dutch and Mandarin: in Dutch, the mass/count distinction is marked through plural morphemes and articles, while in Mandarin, it is marked in the classifier system. Secondly, they need to know that generally there are two groups of classifiers, the count-classifiers and massifiers. They have different functions, lexical meanings, and restrictions. Also, they need to be aware that there is a group of specific classifiers which could have either count-classifier meanings or massifier meanings. Further, they need to know the diagnoses of massifiers: only massifiers can be modified by adjectives, and can be followed by the modification marker *de*, but not count-classifiers.

To sum up, Dutch-Mandarin learners' acquisition of the count/mass distinction in Mandarin involves three missions: 1). Learning that the count/mass distinction is encoded in the classifier system; 2). Learning that there are count-classifiers as well as massifiers, and they have different functions and different abilities of occurring with varied types of nouns (the lexical knowledge of classifiers); 3). Learning that count-classifiers and massifiers have different distributions, they differ in the ability of being modified by adjectives, and being inserted with the modification marker *de* (the syntactic knowledge of classifiers).

Since Dutch-Mandarin learners do not need to learn the mass/count category, instead, they need to learn a new way of expressing the mass/count distinction. In other words, what they need to acquire are the new feature values in Mandarin. Along with the Feature Reassembly Hypothesis (Ladiere, 2009) discussed in Chapter 1, I predict that high

proficiency Dutch-Mandarin learners can reassemble these new features in Mandarin, and exhibit native-like patterns in on-line predictive processing.

As mentioned in Chapter 1, one of the purposes of the current research is to investigate whether implicit learning of unique-to-L2 knowledge could happen on late Dutch-Mandarin learners after a prolonged immersive experience in a Mandarin-dominant environment. The experimental materials in the current research need to be ‘unique’ as well as ‘implicit’. The ‘uniqueness’ of the classifier system and the mass/count syntactic cues in Mandarin have just been presented and validated in Section 1 and Section 2. In the next section, the ‘implicitness’ of the mass/count syntactic cues is discussed.

### **3. Mass/count structures are implicit**

Different researchers have different definitions of implicit and explicit knowledge. According to R. Ellis (2005), implicit knowledge is intuitive and tacit and can be automatically activated during real-time processing without awareness, while explicit knowledge is conscious-involved and usually can generate verbalizable rules. Based on Williams, J. N. (2009), implicit knowledge can be used in the absence of awareness while explicit knowledge needs effort and consciousness to be activated. Based on previous studies, the common characters of implicit knowledge can be summarized as ‘unconscious’, ‘automatic’, and ‘non-verbalizable rules’.

In Mandarin Chinese, apart from the obligatory presence of a classifier between a numeral and a noun, and some high frequency fixed classifier-noun pairs, the differences between count-classifiers and massifiers (e.g. the distributional differences, the differences of the restrictions of classifier-noun pairs) and the mass/count syntactic cues are implicit to native Mandarin speakers. Native Mandarin speakers automatically produce and comprehend nominal phrases with either count-classifiers or massifiers. This is done without being aware of the relations between the nominal phrase structures (i.e. word order) and the mass/count interpretations of the classifiers and nominal phrases. They cannot explicitly verbalize the hidden rules governing the usage of count-classifiers and massifiers. As for L2-Mandarin learners, no textbook or any formal instructions delivered in Chinese-teaching classes by professional teachers includes the difference



between count-classifiers and massifiers, and the mass/count syntactic cues. High proficiency L2-Mandarin learners have been taught the obligatory appearance of a classifier between a numeral and a noun, and some high-frequency classifier-noun pairings, but not the difference between mass/count classifiers and the mass/count syntactic cues in Mandarin (based on the current structure of Hanyu Shuiping Kaoshi since 2010, Official HSK Centre Introduction, HSK Centre; the Official Chinese teaching text book – Edexcel GCSE Chinese Student Book).

Further, to make sure that the mass/count relevant information expressed through classifiers can only be acquired implicitly through immersive Mandarin-dominant environment, but not taught in classrooms, a survey was conducted on Chinese-teachers in Confucius Institution. Twenty Chinese teachers from Confucius Institution in UCL, Queen Mary University of London, SOAS University of London, and Leiden University were participated in this on-line survey. There are two parts in the survey. The first one is a grammar test, in which Chinese-teachers were asked to judge whether a nominal phrase is grammatical or not based on a seven-point scale (1 stands for ungrammatical while 7 stands for grammatical). Every nominal phrase has four different structures, as illustrated in (28).

- (28). a. Num+Cl+Adj+N  
b. Num+Adj+Cl+N  
c. Num+Cl+*de*+Adj+N  
d. Num+Adj+Cl+*de*+N

Both typical count-classifiers (e.g. *ge*) and massifiers (e.g. *bei*), and appreciate nouns were used. The results showed that all Chinese-teachers treated the nominal phrases with the structure in (22b) and (22c) as ungrammatical when the classifier was a typical count-classifier (the mean scores for these two conditions are significantly lower than the other two conditions,  $ps < 0.001$ ; and also significantly lower than when the classifier was a massifier,  $ps < 0.001$ ). This result indicates that these Chinese-teachers have fully mastered the mass/count distinction marked in the classifiers and the mass/count syntactic cues in Mandarin.

In the second part, Chinese-teachers were asked to answer some questions. The questions are listed in (29).

(29). a. Do you know there are count-classifiers and massifiers in Mandarin Chinese? If you do, could you please write down the difference between these two kinds of classifiers, and also give some examples of each kind?

b. Is there any instruction in Chinese-teaching textbooks about count-classifiers and massifiers in Mandarin Chinese? If there is, could you please write down to which level of Chinese-learning students this knowledge is delivered?

c. Have you ever given any instruction in Chinese-teaching classes about count-classifiers and massifiers in Mandarin? If you have, could you please write down to which level of Chinese-learning students? And what was the instruction about?

d. Has any Chinese-learning students asked some questions relevant to count-classifiers and massifiers in Mandarin? If there is, can you please write down what were the questions and how did you answer them?

None of the 20 Chinese-teachers answered ‘yes’ to the questions in (29a), (29b), and (29c). This result convincingly indicates that these Chinese-teachers have no explicit knowledge of the count/mass classifiers, and no experience of teaching this knowledge to L2ers.

In general, the results of the two parts in this survey showed that as native Mandarin speakers, these Chinese-teachers have very little awareness of the count/mass distinction in the classifier system in Mandarin, even though they could use count-classifiers and massifiers automatically and unconsciously. They have no experience of teaching or encountering any instruction about count/mass-classifiers. Some Chinese-learning students have asked some questions about the classifier-noun pairing restrictions, but none of them mentioned the difference between count-classifiers and massifiers.

Based on the analysis of these 20 Chinese-teachers’ results, it is confirmed that native Mandarin speakers, as well as L2-Mandarin learners who have numeral languages (e.g. English and Dutch) as their first languages, can only implicitly acquire the mass/count distinction in the classifier system through immersive daily Mandarin-dominant linguistic input.

To sum up, the mass/count distinction in classifiers and the mass/count syntactic cues are implicit to both native Mandarin speakers and late Dutch-Mandarin learners. In the current research, using dual-role classifiers and nominal phrases with different mass/count syntactic cues, we could gain some understandings about how native Mandarin speakers and high proficiency late Dutch-Mandarin learners on-line process unique-to-L2 knowledge which can only be acquired through implicit learning.

Among several experimental techniques which can be used to test participants' automatic on-line processing of implicit knowledge, the eye-tracking technique is a famous one (Godfroid, Winke, & Gass, 2013). Eye-tracking data are people's eye movements during their reading. It is often used to investigate readers' automatic and unconscious language processing (Godfroid & Winke, 2015; Sagarra & Seibert Hanson, 2011). In this current research, in order to investigate how native Mandarin speakers and late Dutch-Mandarin learners on-line process nominal phrases with different mass/count syntactic cues, the eye-tracking technique is used in two experiments. Detailed information about the research methods used in the current research and previous relevant studies are presented in Chapter 3.

### **III. Previous Studies and Research Methods**

In order to investigate whether high proficiency late Dutch-Mandarin learners can use mass/count syntactic cues in on-line sentence processing and anticipation-building, research methods which can offer time-sensitive and space-sensitive data, and can be used to investigate participants' real time processing were chosen in the current research. In this chapter, the detailed information about the research methods used in the current research is introduced: the Visual World Paradigm and the Reading for Comprehension task. The difference of the required processing loads between these two paradigms is also discussed. Following the discussions in Chapter 2 about the mass/count distinction in Mandarin, previous studies focused on native Mandarin speakers' and L2-Mandarin-learners' processing of classifiers and the mass/count distinctions are reviewed first.

This chapter is structured as follows. In Section 1, previous studies focused on native Mandarin speakers' and L2-Mandarin-learners' processing of classifiers and the mass/count distinctions are reported. In Section 2, the research methods employed in the current research are introduced, and the differences between these two experimental paradigms are discussed.

#### **1. Previous studies**

Some experimental studies have been conducted to investigate Chinese children's and adults' comprehension and use of count-classifiers and massifiers. Considering the classifier system's uniqueness to Mandarin for some L2-Mandarin learners whose native language does not contain the classifier system, some SLA researchers are interested in how L2-Mandarin learners take advantage of the classifier information during sentence comprehension and processing. In this section, previous studies that explored native Mandarin speakers' comprehension and use of classifiers are reported first, followed by

the review of previous studies considering L2ers' use of the classifier information in Mandarin.

### **1.1. Classifier studies on native Mandarin speakers**

Many studies have been conducted regarding Chinese-speaking children's acquisition of classifiers (Chang, 1983; Fang, 1985; Hu, 1993a, 1993b; Ying, Chen, Song, Shao, & Guo, 1983). Most of them focused on count-classifiers only, except for the study conducted by Ying et al. (1983) which tested the acquisition order of both count-classifiers and massifiers. These previous studies explored Chinese children's acquisition of classifiers, but left the question open whether children acquire knowledge of the mass/count distinction in their early stages of language acquisition, something which has been extensively investigated in studies on English-speaking children (Brown, 1973; Soja, Carey, & Spelke, 1991).

Chien, Lust, & Chiang (2003) is the first experimental study testing Mandarin children's comprehension of the mass/count classifiers. Eighty Mandarin-speaking children (aged between 3 and 8 years old) and 16 Mandarin-speaking adults were recruited as participants. Each participant was required to choose one out of three items based on the following instruction: 'Mickey Mouse wants one *classifier* (CL) X' (where X is a novel word). Either a count-classifier or a massifier was embedded in Mickey Mouse's request, creating either a count-selective context or a mass-selective context. They predicted that if a child possesses knowledge of the grammatical mass/count distinction, he or she should predominantly use the count-selective context (i.e. the count-classifier) to choose an entity corresponding to a count noun, and the mass-selective context (i.e. the massifier) to choose an entity corresponding to a mass noun. They found that Mandarin-speaking children are sensitive to the mass/count classifiers even at early stages of language acquisition (as young as 3-years-old). Their performance on the count-classifier-noun pairs and the massifier-noun pairs are comparable. These results are consistent with the claim from Cheng & Sybesma (1998, 1999) that the mass/count distinction is grammatically encoded in the classifier level in Chinese. The observations in this study indicated that Mandarin-speaking children (as young as 3-years-old) already possess the distinction of mass/count classifiers and mass/count object representations. They can map these two kinds of information to each other during sentence comprehension. The results implied that the mass/count semantic information embedded in mass/count

classifiers can be used to predict or interpret upcoming nouns for Mandarin-speaking children older than 3 years old.

To further explore native Mandarin speakers' use of mass/count semantics as well as mass/count syntax, Li, Barner, & Huang (2008) conducted three off-line tests. In the first two experiments, similar to Chien et al. (2003), they tested whether count-classifiers cause participants to expect solid individual objects while massifiers select non-solid substance. Three boxes were presented to each participant and each box contained an item. These three boxes were: an open box containing an unfamiliar solid object, an open box containing a group of unfamiliar non-solid substance which was shaped in a certain way, and a closed box. The experimenter asked for one of the items (e.g. '*I want one classifier X*'. X in the instruction is a novel word, so the only useful information in the instruction is the classifier.). Participants were required to select a box based on the instruction. There were three types of trials in Experiment 1: the no-shape match trial, the solid shape match trial, and the nonsolid-shape match trial. Only count-classifiers were used in Experiment 1. They predicted that if participants are sensitive to the selective properties of the classifiers, they should choose different boxes in different trials. In the no-shape match trial, participants were expected to choose the closed box since the shape of neither object nor substance matched the shape described by the classifier. In the solid shape match trial, participants were expected to choose the open box with one unfamiliar solid object since they heard a count-classifier from the experimenter which should select solid objects. While in the nonsolid-shape match trial, if count-classifiers select count solid objects, participants were expected to choose the closed box when they heard a count-classifier even if the unfamiliar non-solid substance was shaped in the way the classifier depicted. The results showed that adults Mandarin speakers behaved exactly as predicted. The results indicated that native Mandarin speakers are sensitive to the semantic properties of mass/count classifiers: on hearing count-classifiers, they were deliberately looking for solid objects and ignoring non-solid substances.

In Experiment 2, they used both count-classifiers and massifiers in the instruction. For count-classifiers, they expected participants to pick the closed box in the no-shape match trial, pick the solid object in the shape match trial, and pick the closed box in the nonsolid-shape match trial. For massifiers, participants were expected to pick the closed box in the no-shape match trial and the shape match trial, and pick the non-solid

substance in the nonsolid-shape match trial. Differing from Experiment 1, there was an additional trial in Experiment 2: the multiple solids trial. In this trial, there were three boxes: an open box with multiple solids organized in a certain way which is consistent with the classifier, an open box with nonsolid substance organized in a different shape, and a closed box. Participants were expected to pick the open box with multiple solids when they heard a massifier from the experimenter since massifiers select a mass or plural set meaning. They were expected to pick the closed box if they heard a count-classifier. Again, the results supported their predictions.

The results of these two experiments showed that Mandarin speakers are sensitive to the semantic information of classifiers. There is a semantic mass/count distinction on the classifier-noun pairings: when the classifier is a count-classifier, participants prefer to choose solid objects rather than non-solid substance, even when the non-solid substance is shaped in a way that is consistent with the classifier. When the classifier is a massifier, participants choose the box containing either multiple solid objects or non-solid substance when the way these items are organized is consistent with the classifier. These findings indicated that in Mandarin Chinese, count-classifiers select solid objects while massifiers select non-solid substance and multiple objects. This is consistent with the discussion in Chapter 2 that, count-classifiers usually occur with count nouns with individual readings, while massifiers occur with both mass nouns and count nouns with plural/portion readings.

To further explore the influence of the mass/count syntax on native Mandarin speakers, in Experiment 3, Li et al tested Mandarin speakers' sensitivity to 'mass/count sensitive structures'. As discussed in Chapter 2, according to Cheng & Sybesma (1998, 1999), only massifiers can be modified by adjectives, and the modification marker *de* can only optionally occur after a massifier. The 'mass/count sensitive structures' included two types: [Num+Cl+Adj+N], and [Num+Adj+Cl+*de*+N]. For each participant, there were two choices: one whole object, or a part/group of the objects (e.g. a whole CD vs. a broken portion of a CD, or an individual fan vs. a bunch of fans). Participants were asked to match these choices with the two types of nominal phrases: [one CL small N] and [one small CL *de* N]. An example of the materials is illustrated in (30).

(30). An example from Li et al. (2008)



Li et al. (2008) argued that, if the claim in Cheng & Sybesma (1998, 1999) is valid, it should be expected that putting a count-classifier in a nominal phrase with the sequence [Num+Adj+Cl+*de*+N] would force readers to have a coerced mass/plural interpretation of the object, just like English count nouns used in a mass syntax structure: 'there is dog all over the wall' (Universal grinder effect, See Pelletier, 1975; Cheng, Doetijes, & Sybesma, 2008). They further predicted that if Mandarin speakers are sensitive to this mass/count syntax, different structures will lead to different behaviours: for the nominal phrases with the sequence [one CL small N], participants would choose one whole object, while for nominal phrases with the sequence [one small CL *de* N], participants would choose a part/group of objects. Four classifiers were used in Experiment 3: *gen* 'rod', *kuai* 'chunk', *pian* 'slice', and *ba* 'handful', which are all count-classifiers according to Li et al. (2008).

The results showed that adult native Mandarin participants were sensitive to the mass/count syntax: they chose the countable objects with individual discrete units (e.g. a CD, a fan) when heard [one CL small N], and chose a part/group of the objects (e.g. a part of a CD, a bunch of fans) when heard [one small CL *de* N]. Based on these observations, the authors argued that different nominal phrase structures affect Mandarin speakers' interpretations of nominal phrases: when classifiers are count-classifiers, they prefer to interpret the nominal phrase with the structure [Num+CL+Adj+N] as countable objects with discrete units, while interpreting nominal phrases with the structure [Num+Adj+CL+*de*+N] as substances or objects shaped/organized in a certain way.

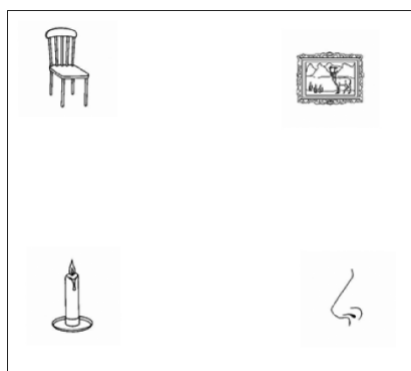
To sum up, Li, Barner & Huang (2008) offered experimental evidence that native Mandarin speakers are sensitive to mass/count semantics: count-classifiers occur with



solid individual objects while massifiers occur with non-solid substances and multiple objects. Furthermore, native Mandarin speakers are sensitive to mass/count syntax: compared to the structure [Num+CL+Adj+N], the structure [Num+Adj+CL+*de*+N] forces a mass/plural interpretation.

Recently, more and more studies are starting to use on-line processing paradigms to explore native Mandarin speakers' on-line processing of classifiers and classifier-noun pairs. Huettig, Chen, Bowerman, and Majid (2010) used a Visual World Paradigm in two experiments to explore whether classifier categories can be recruited during online processing by native Mandarin speakers. Simple Mandarin sentences were played through headphones while participants' eye movements to four pictures presented on a screen were monitored. In each trial, one of the four pictures represents the object sharing the same classifier with the target noun in the audio sentence, the other three pictures represent irrelevant objects (distractors). The researchers mainly concerned about what picture participants were looking at when hearing the target noun embedded in the sentence. They predicted that if classifier categories affect general conceptual processing, on hearing the target noun, participants should shift their attention to objects that are members of the same classifier category even when the classifier is not overtly presented. For instance, when hearing the target noun *yaoshi* 'key', participants should land more fixations on a picture of a chair than other irrelevant pictures since *yaoshi* 'key' and *yizi* 'chair' share the same classifier *ba* in Mandarin. In Experiment 1, the classifier was spoken overtly (with the audio sentence 'Do you know if there is another name for one CL key?'). While in Experiment 2, there was no overt classifier presented (with the audio sentence 'I am looking for a word that rhymes with key.'). An example of the materials is illustrated in (31).

(31). An example from Huettig et al. (2010)



The results showed that participants shifted significantly more fixations to the classifier-relevant pictures (e.g. the chair) than the irrelevant distractors in Experiment 1 where the classifier was explicitly presented. In Experiment 2, however, no such classifier category effect was observed. They argued that the results suggested that the classifier information affects native speakers' fixation behaviours only when it is necessary to process the classifier information during linguistic processing. Also, in Experiment 1, native speakers started to land more fixations on the classifier-relevant picture on hearing the classifier, even before they reached the target noun. This observation indicated that during on-line processing of Mandarin sentences, classifiers can be treated as a predictive cue for listeners to anticipate the classifier-associated nouns to be the upcoming noun. In other words, for native Mandarin speakers, the processing of a classifier can pre-activate the processing of its associated nouns. Native Mandarin speakers can use classifier information to build anticipations for the upcoming nouns in real-time processing.

Based on Huettig et al. (2010), Klein, Carlson, Li, Jaeger, & Tanenhaus (2012) conducted two Visual World Paradigm experiments on native Mandarin speakers to further investigate how massifiers (Experiment 2) and count-classifiers (Experiment 3) affect native speakers' on-line processing and comprehension. Participants were asked to pick one out of four pictures based on the instructions they heard from the headphones. There were two conditions in each experiment: a general condition and a specific condition. In the general condition, a very general massifier/count-classifier was used which could occur with all the four entities presented by the four pictures, and thus cannot offer any specific information about the upcoming noun (e.g. '*Choose a PICTURE of lettuce*'). In the specific condition, a very specific massifier/count-classifier was used which could be treated as an informative instruction since it can only occur

with the target object and its classifier competitor (e.g. ‘Choose a *HEAD* of lettuce’). The four pictures in each trial were: a group of target objects, a group of objects which are phonological competitors of the target object, a group of objects which share the same massifier/count-classifier with the target object, and a group of objects which are phonological competitors of the massifier/count-classifier. In the audio sentences, the general massifier was *xie* ‘some’, the general count-classifier was *ge* ‘unit’, the specific massifier was *shu* ‘bunch’, and the specific count-classifier was *shan* (the classifier used with ‘door’, ‘window’, or something big and flat). An example of the materials is shown in (32).

(32). An example from Klein et al. (2012)



Participants were expected to exhibit a phonological cohort effect in the general condition since the general massifier/count-classifier cannot offer any information about the upcoming noun. Thus they need to wait until the onset of the noun to select the right picture. At this point, participants would begin to look at the two similar-sounding pictures shortly after the onset of the noun and exhibit a phonological cohort effect. For example, the target *baihe* ‘lily’ and its phonological cohort *baicai* ‘cabbage’ should compete for the selection. In the specific condition, however, participants were expected to look at the two pictures which share the same specific massifier/count-classifier after the onset of the classifier since the specific massifier/count-classifier offers information about the upcoming nouns. For example, when the massifier was *shu* ‘bunch’, participants should fix more on the two pictures which contain ‘three bunches of lilies’ and ‘three bunches of rose’ after the onset of *shu* ‘bunch’ than the other two pictures which cannot occur with the specific massifier. The results showed that the target picture receives most of the fixations about 100 ms quicker in the specific condition than in the general condition in both experiments. This advantage demonstrated that in Mandarin,

native speakers can make a swift use of the classifier information during on-line processing to build anticipations for the upcoming nouns no matter whether the classifier is a count-classifier or a massifier. Similar to the findings in Huettigs et al. (2010), the results of Klein et al. (2012) indicated that in Mandarin, classifiers can be used by native speakers in prediction forming and on-line processing to narrow down the possible options.

In general, from above discussions we can see that both off-line tests (the forced choice task, Chien et al., 2003; Li et al., 2008) and on-line tasks (the Visual World Paradigm, Huettig et al., 2010; Klein et al., 2012) have been used to explore native Mandarin speakers' understanding and processing of nominal phrases with different classifiers. There are two main findings from these previous studies: 1) different nominal phrase structures give rise to different interpretations (Li et al., 2008); 2) the information encoded in classifier can be used during on-line processing to build anticipations about the upcoming nouns by native Mandarin speakers (e.g. Huettig et al., 2010; Klein et al., 2012). In this current research, I expect native Mandarin participants to display their sensitivity to the mass/count structures (mass/count syntactic cues) as well as the ability to use the classifier information to build anticipations during on-line processing.

### **1.2. Classifier studies on L2-Mandarin learners**

As unique-to-classifier-languages constructions, the classifier system and classifier-noun pairs in Mandarin Chinese have attracted the attention of some SLA researchers. Some studies have been conducted to explore the situations of L2ers' acquisition and processing of classifiers in Mandarin Chinese.

Polio (1994) investigated the influence of language distances on late L2ers' usage of classifiers in Chinese. Twenty-one English-Chinese learners and 21 Japanese-Chinese learners were recruited. She predicted that if language distance affects late L2ers' acquisition and production, it should be easier for Japanese-Chinese learners to produce classifiers in a native-like way than English-Chinese learners, since both Japanese and Chinese are considered as classifier-languages. A story-retelling paradigm was used in her research, in which a silent film was showed to participants. Participants were required to retell the story to a native speaker of Chinese, and their speech was recorded and analysed. The results showed that both the two groups of L2ers rarely omitted classifiers when producing nominal phrases. The obligatory presence of a classifier between a

numeral and a noun has been well acquired by both groups of participants. However, while they were highly sensitive to the syntactic positions of classifiers, they were less sensitive to the semantic properties: L2ers primarily used the general classifier *ge* in most situations instead of the appropriate specific classifiers. This behaviour pattern has also been found in several studies on Mandarin-speaking children. Fang (1985) and Hu (1993a) found that Mandarin-speaking children first acquire the general classifier *ge* and use it as a ‘syntactic place-holder’: they overuse the general classifier *ge* and allow it to co-occur with almost any noun which in adult language is considered ‘informal’ and ‘inappropriate’. Comparing the two groups of L2ers Polio found that compared to English-Chinese learners, Japanese-Chinese learners exhibited some beneficial transfer from their L1, but also some negative influences on cases where Japanese and Chinese share the same character but with different semantic meanings. In general, using a story-retelling paradigm Polio (1994) found that in Chinese production, L2ers were aware of the obligatory presence of a classifier in a numeral phrase. They preferred to use the general classifier *ge* in most cases. Like Chinese children, late L2ers developed the syntax of classifiers first, and the lexical-semantics later. Cross-language similarity has some effects on L2ers’ classifier production, which can be either positive transfer or negative transfer.

Liang (2009) conducted three off-line tests to explore L2ers’ acquisition of various classifiers: a classifier comprehension test, a classifier production test, and a classifier prototype test. In the classifier comprehension test, participants were asked to match objects to classifiers in a list. In the classifier production test, participants were asked to produce a classifier which could occur with the objects presented in pictures. In the classifier prototype test, participants were asked to rank the possible nouns which could occur with the given classifier by their prototypicalities. Twenty-nine English-Chinese learners and 29 Korean-Chinese learners with different Chinese proficiency levels were recruited. The results showed that the correlation between L2ers’ performance and their Chinese proficiency level is positive: the higher the proficiency, the more native-like behaviour the participants exhibited. Also, similar to Polio (1994), Liang found that language distance has an effect on L2ers’ classifier acquisition: in general, Korean-Chinese learners outperformed English-Chinese learners.

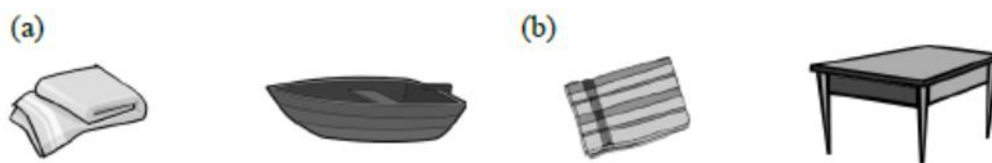
Both Polio (1994) and Liang (2008) used only count-classifiers in their research and failed to take into consideration massifiers, without which it would be difficult to get a

complete picture of L2ers' classifier processing and acquisition. Gong (2010) added massifiers in his materials to fill this gap. Nine English-Chinese learners were tested on their production of both count-classifiers and massifiers. The results illustrated similar patterns as those in Polio (1994): L2ers overgeneralized the general classifier *ge* in a way similar to Chinese children. Apart from that, Gong (2010) also found that L2ers' acquisitions of count-classifiers and massifiers are comparable.

Taken together, previous studies using off-line tests on L2-Chinese learners found that L2ers are aware of the syntactic requirement of classifiers, since they always use a classifier (usually the general classifier *ge*) to fill the syntactic classifier position. However, compared to syntactic information of classifiers, lexical-semantic information of classifiers and the semantic restrictions of classifier-noun pairs are more difficult to be acquired: L2ers often choose the general classifier *ge* and ignore the semantic constraints between classifiers and their associated nouns.

Recently, some researchers have started to use some on-line processing paradigms to investigate L2ers' processing of classifiers in real-time. Lau & Grüter (2015) conducted a Visual World Paradigm experiment to investigate the predictive using of classifiers during on-line processing on L1 and L2 speakers of Chinese. Both native Mandarin speakers and English-Chinese learners who were at beginner to intermediate Chinese proficiency levels were recruited as participants. Participants were required to listen to the instructions from the headphones and choose one from the two pictures presented on the screen based on the instruction. Two classifiers were used in their study: *tiao* (usually occurs with long, string-like objects) and *zhang* (occurs with flat surfaced objects). The audio materials had two conditions: the SAME condition in which the two pictures presented on the screen share the same classifier, and the DIFFERENT condition in which the two pictures have different classifiers. An example of the materials is in (33).

(33). An example from Lau et al. (2015)



They predicted that if participants are sensitive to the classifiers' predictive information, they should fix on the target picture faster in the DIFFERENT condition than in the SAME condition. Since in the DIFFERENT condition the two pictures have different classifiers, the classifier in the audio instruction could be treated as an informative cue which helps participants build anticipations for the associated nouns to be the upcoming noun, and thus narrow down the options. On the other hand, in the SAME condition, the two pictures share the same classifier, the classifier in the audio instruction would not be informative since it fits with both nouns denoted by the two pictures. Participants need to wait till the onset of the noun to decide which is the target picture.

The results showed that native Mandarin speakers behaved exactly as predicted: they fixed on the target picture faster in the DIFFERENT condition than in the SAME condition. L2ers were in general slower than native speakers, and they were slightly but significantly faster in the DIFFERENT condition than in the SAME condition. These findings indicated that for both L1ers and L2ers, classifier information can be used during on-line processing, and the processing of a classifier could prime or pre-activate the processing of its associated nouns.

To conclude, previous studies on L2-Chinese learners found that: 1) there is a positive relation between L2ers' performance on the processing and understanding of classifiers and their Chinese proficiency levels, that is, higher proficiency L2ers always exhibited better performance; 2) L2ers are sensitive to the syntactic place of classifiers, but not the semantic connections between classifiers and nouns; 3) similar to native Mandarin speakers, L2ers can take advantage of the classifier information during on-line processing to build anticipations for the upcoming nouns. Based on these existing findings, the high proficiency Dutch-Mandarin learners in this current research are expected to exhibit native-like patterns in the predictive using of classifiers during on-line processing.

To sum up, previous research about the processing of nominal phrases by both native Mandarin speakers and L2-Mandarin learners can be divided into different categories based on the manipulations of three main factors: materials, research methods, and participants. In general, there are two levels of the factor 'materials': studies have been conducted either on different classifiers (Chien et al., 2003; Gong, 2010; Liang, 2008,

etc.), or on different nominal phrase structures (Li, Barner, & Huang, 2008). The factor ‘research methods’ also has two levels: studies used either off-line tests (Chien et al., 2003; Gong, 2010; Liang, 2008, etc.), or on-line tasks (Huettig et al., 2010; Klein et al., 2012, etc.). The factor ‘participants’ has two levels: studies focused on native Mandarin speakers (Chien et al., 2003; Huettig et al., 2010, etc.) or L2-Chinese learners (Gong, 2010; Liang, 2008; Lau & Gruter, 2015, etc.). From an experimental research perspective, any relevant research questions can be realized through manipulating these three main factors. Categorizing previous studies through these three factors, I provide a summary in Table 6.

*Table 6. A summary of previous studies*

<b>Materials</b>	<b>Research method</b>	<b>Participants</b>	<b>Previous studies</b>
Different classifiers	Off-line tests	L1ers	Chien, Lust, & Chiang (2003)
Different classifiers	Off-line tests	L2ers	Gong (2010); Liang (2008); etc.
Different classifiers	On-line tasks	L1ers	Huettig et al. (2010); Klein et al. (2012)
Different classifiers	On-line tasks	L2ers	Lau & Gruter (2015)
Different structures	Off-line tests	L1ers	Li, Barner, & Huang (2008)
Different structures	Off-line tests	L2ers	-
Different structures	On-line tasks	L1ers	<b>The current research</b>
Different structures	On-line tasks	L2ers	

From Table 6 we can see that most of the previous studies were using either on-line tasks or off-line tests to investigate L1ers’ and L2ers’ acquisition and processing of nominal phrases with different classifiers. There is only one study (Li, Barner, & Huang, 2008) which explored the influence of different nominal phrase structures on native Mandarin speakers’ comprehension of nominal phrases.

Despite their contributions to the findings of native Mandarin speakers’ sensitivity to the mass/count syntactic cues, there are some improvements that can be done based on Li et



al. (2008)'s research if we take a close look at their materials. First of all, the 'count-classifiers' they used in Experiment 3 are not count-classifiers according to Cheng (2011, 2012). The four classifiers in Experiment 3 were *gen* 'rod', *kuai* 'chunk', *pian* 'slice', and *ba* 'handful', which are all massifiers if we adapt the categories in Cheng (2011, 2012). From this point of view, their results did not argue that putting a count-classifier in a nominal phrase with the massifier structure [Num+Adj+CL+*de*+N] would force participants to have a coerced quantification/mass interpretation of the nominal phrase. Instead, the results suggested that putting these classifiers in nominal phrases with different mass/count syntactic cues would lead to different interpretations.

Secondly, the materials in Li et al. (2008) could be more strictly controlled. They used nominal phrases with two structures in Experiment 3: the neutral structure [Num+CL+Adj+N] and the massifier structure [Num+Adj+CL+*de*+N]. According to Cheng & Sybesma (1998), the co-occurrence with the modification marker *de* and the classifiers' possibility of being modified by adjectives like 'small' and 'big' are two syntactic cues which could be helpful to tell count-classifiers and massifiers apart. Li et al. (2008) argued that nominal phrases with the massifier structure [Num+Adj+CL+*de*+N] force participants to have a mass/plural interpretation. But what remains unclear is which part of the massifier structure [Num+Adj+CL+*de*+N] forced native Mandarin speakers to have a mass/plural interpretation. Is it because of the modification of the classifier, or the insertion of the modification marker *de*, or both?

Furthermore, the paradigms Li et al (2008) used in their three experiments were off-line forced choice tasks. Participants have plenty of time and may use explicit task-dependent strategies to parse what they had heard and make decisions about what reactions they should take. In order to get a clear picture about how Mandarin speakers and L2-Mandarin learners react to the 'mass/count sensitive structures' in real-time, and whether they have acquired the mass/count syntactic cues as implicit knowledge, it is important to use on-line processing paradigms and recording techniques with high temporal resolution to measure participants' processing patterns.

Moreover, the critical key words in their materials were novel words (in Experiment 1 and Experiment 2), and the objects were unknown items (in Experiment 3). It would be interesting to examine how native Mandarin speakers and L2-Mandarin-learners react

when encountering real words and common objects in nominal phrases with different syntactic cues.

In order to fill these gaps and reach a better understanding of how different nominal phrase structures affect native Mandarin speakers' and Dutch-Mandarin learners' on-line processing, the current research used two eye-tracking experiments to explore both groups' on-line processing and interpretation of nominal phrases with different mass/count syntactic cues. In this current study, the potential effect of classifiers' possibility of being modified by adjectives was tested in the Visual World Paradigm experiment, while both the insertion of the modification marker *de* and the ability to be modified by adjectives were manipulated in the Reading for Comprehension experiment. In Section 2, research methods used in the current research are introduced and discussed.

## **2. Research methods**

Eye-tracking techniques have been used to investigate cognitive processing since Javal (1879)'s first study on eye movements in reading (see a review of Huettig, Rommers, & Meyer, 2011). Boosted by remarkable improvements in eye-tracking recording systems since 1970s, and huge developments of general theories of language acquisition and sentence processing, in recent years, more and more eye-tracking studies have been conducted to test on-line language processing and cognitive processing (Keating, 2009; Lim & Christianson, 2014; N. Ellis, Hafeez, Martin, Chen, Boland, & Sagarra, 2014; Sagarra & N. Ellis, 2013; also see a review of Clifton Jr., Ferreira, Henderson, Inhoff, Liversedge, Reichle, Schotter, 2016). Sentence on-line processing reflects participants' automatic use of implicit knowledge without awareness (R. Ellis, 2005; Jiang, 2007), which can be revealed by participants' eye movements (Rayner, 1998, 2009). And further, processing difficulties can be captured in eye movements records (Rayner, Reichle, & Pollatsek, 2005; Reichle, Rayner, & Pollatsek, 2003). Thus in this current research, the eye-tracking technique is used to investigate native Mandarin speakers' and late Dutch-Mandarin learners' on-line processing and anticipation building.

Among several eye-tracking experimental paradigms (e.g. moving window paradigm, moving mask paradigm), the Visual World Paradigm and the Reading for

Comprehension task were chosen in the current research. In this section, detailed information about these two paradigms are introduced, and the differences between them are discussed.

### **2.1. Visual World Paradigm**

The Visual world paradigm was first used by Cooper (1974). In his study, Cooper found that when participants were asked to listen to short narratives while looking at displays showing objects, they fixed more on the objects which were referred to in the spoken text than other irrelative objects. Also, he found that participants' eye movements were time-sensitive to the audio material: most of the fixations on the related objects were triggered either while the corresponding word was spoken or within 200 ms after the word offset. Based on these observations, Cooper claimed that there exists a systematic relationship between eye movements and language processing, and this Visual World Paradigm could be used to explore listeners' real-time perceptual and cognitive processing of sentences. Recently, the Visual World Paradigm is more often used to investigate participants' anticipatory effects, especially when they launch their eye fixations to a certain region of the screen (i.e. the target picture or object) before the name of the picture/object has actually been pronounced in the audio materials (Hopp, 2012, 2015; Tanenhaus & Trueswell, 2006; Trenkic, Mirkovic, & Altmann, 2014). The change of participants' fixations in different regions along the input of the audio materials reveals their real-time integrations of different sources of input (e.g. lexical, semantic, syntactic, contextual). This index the level of activation of the different audio input candidates, their competition, and the ultimate selection of a certain input as the target (Winke, Godfroid, & Gass, 2013). Furthermore, since anticipatory processing is an automatic and unconscious procedure, the Visual World Paradigm can thus reveal participants' on-line use of implicit knowledge (Godfroid & Winke, 2015). In the current research, the Visual World Paradigm was used to test native Mandarin speakers' and Dutch-Mandarin learners' predictive using of unique-to-L2 constructions (i.e. the mass/count syntactic cues) which can only be acquired through implicit learning.

The basic setup of the Visual World Paradigm is simple. Participants are required to listen to some speech streams from headphones while looking at displays of objects or pictures on the computer screen. Some of the objects or pictures are mentioned in the spoken utterance (targets), while others are not. Computer and eye-tracking software record participants' eye movements while they carry out the task based on the

instructions from the headphones (Huettig & McQueen, 2007; McQueen & Viebahn, 2007). The change of the percentage of participants' fixations in each picture/object, or the target picture/object along the audio input is the main concern in the Visual World Paradigm.

The most powerful contribution of the Visual World Paradigm is its ability to estimate the fine-grained time courses of activation and competition among visual tasks during linguistic input processing. The properties of the pictures and the linguistic inputs create an underlying probability distribution of fixation locations, and the observed fixation proportions reflect this underlying distribution. In this paradigm the main focus is how the proportions of fixations in a certain region (picture) change over time (with different linguistic inputs), and how the manipulated factors (characters of pictures, or different versions of audio inputs) affect this change. The data analyses in the Visual World Paradigm focuses on the question of how likely participants are to look at a specific region of interest at different times (on hearing different audio inputs) in a trial. The most common dependent variable is the fixation proportion on the interest area in each time window (Altmann, 2004). The selection of interest regions depends on varied research questions. Usually the audio input of the corresponding interest region is embedded in a carrier sentence (e.g. 'Please find/choose/click X among the objects on the screen', where X represents the audio input of the interest region), participants thus have a few seconds to familiarize themselves with the objects/pictures prior to the onset of the interest region.

The Visual World Paradigm has been widely used to investigate real-time language processing and anticipation building (Hopp, 2012, 2015; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995; Allopenna, Magnuson, & Tanenhaus, 1998). It has been confirmed by several studies that listeners prefer to look at the objects that are mentioned or implied/predicted by what they heard (Land, Mennie, & Rusted, 1999; Irwin, 2004). Listeners also intend to relate the audio materials they heard to the visual materials to connect these two types of input with each other (Altmann & Kamide, 2007; Malpass & Meyer, 2010). Altmann and Kamide (2007) have proposed that an increase in the activation of a mental representation of an object (e.g. by linguistic input) results in the increased likelihood of a fixation on that object. Further, several studies found that the predictions which are formed based on the audio input cause an increased likelihood of fixations on the target picture (Hopp, 2012, 2015; Trenkic, Mirkovic, & Altmann, 2013),

as have mentioned in Chapter 1. These previous studies offered convincing evidence that the Visual World Paradigm can be used to investigate participants' automatic on-line processing and anticipation building of language input since it looks into participants' tendency to look at certain parts of the visual displays which is caused by the audio input. In the Visual World Paradigm, participants' eye movements (especially their fixations) reflect their real-time cognitive processing and prediction forming which is an unconscious automatic procedure.

As a time-sensitive experimental paradigm, the Visual World Paradigm is an ideal tool to investigate participants' on-line processing and anticipation building. As discussed above in Section 1, previous studies have already showed that the information embedded in classifiers has an effect on participants' on-line processing of nominal phrases (Huettig et al., 2010; Klein et al., 2012). Very rarely has research explored different nominal phrase structures' influence on L1ers' and L2ers' on-line processing and anticipation building. To the best of my knowledge, so far, Li et al. (2008) is the only research which explored different nominal structure's effect on native Mandarin speakers' interpretations of nominal phrases. It still remains unclear how native Mandarin speakers on-line process and interpret nominal phrases with different structures, and how they use the classifier information to build anticipations in real-time. Furthermore, no study has ever investigated whether high proficiency late L2-Chinese learners could use the classifier information to build anticipations automatically and unconsciously. To fill these gaps, in the current study, through the Visual World Paradigm and by looking at the changes of fixation proportions on different pictures over time, one could tell how different nominal phrase structures (i.e. mass/count syntactic cues) affect both native Mandarin speakers' and Dutch-Mandarin learners' on-line processing and anticipation building. The detailed information about the Visual World Paradigm experiment in the current research is reported in Chapter 4.

## **2.2. Reading for comprehension task**

The Reading for Comprehension task is a typical reading task based on eye-tracking techniques (Clifton, Staub, & Rayner, 2007; Rayner, 1998, 2009). Used in language processing, eye-tracking technique supports complex tracing of saccades (eye jumps), fixations (eye stops), gaze time (eye stops before move right forwardly) and regressions (re-takes) while readers read sentences on the screen. All these indexes provide time-

sensitive (millisecond, ab. ms) and space-sensitive (to the letter) information which reflects readers' on-line conceptual and syntactic processing as well as off-line processing. Due to its ability to provide rich-data collection and multi-level information, the Reading for Comprehension task based on eye-tracking has been an important tool for cognitive-psychology researchers and psycholinguistics to investigate how people process sentences on-line and off-line (Blumenfeld & Marian, 2011; Duyck, Van Assche, Drieghe, & Hartsuiker, 2007; Felser, Sato, & Bertenshaw, 2009; Flecken, 2011).

For decades, the widest known research tool to look into the on-line processing of sentences has been the self-paced reading task (especially the non-cumulative technique, Just et al., 1982). In a self-paced reading task, sentences are presented word by word. When participants finish reading the current word, they press a certain button, the current word disappears and is replaced by the next word. Compared to the self-paced reading task, which can only record participants' reaction time and button-pressing response, the Reading for Comprehension task based on the eye-tracking technique can provide multiple indexes (first fixation, gaze duration, total reading time, etc.) (Frenck-Mestre, 2005). In addition, with the eye-tracking technique, one can distinguish first run processing from second run processing, and from all the other later stages of processing. Researchers can look into the early stage of processing as well as the late stage. While in a self-paced reading task, it would be very unlikely to measure first pass indexes and second pass indexes separately (Clifton, Staub, & Rayner, 2007; Rayner, 1998, 2009). Moreover, sentences in the Reading for Comprehension task can be presented in a natural condition, while in the self-paced reading task, they need to be broken into segments (Jiang, 2004; Keating, 2013). Taken together, the Reading for Comprehension task based on the eye-tracking technique is more time-sensitive and a richer information source in investigating on-line processing compared to the self-paced reading task (for reviews, see Frenck-Mestre, 2005). Considering the main concern in the current research is looking into participants' early stage processing as well as late stage processing, the Reading for Comprehension task was employed.

The basic set-up in the Reading for Comprehension task is simple and straightforward: participants are required to keep their head still during the whole experiment, they can only accomplish the reading task by moving their eyes. Sentences are presented on the screen. After each sentence, there is a meaning-related question presented on the screen. Participants are asked to read the sentences silently for comprehension, and answer the

question after each sentence by pressing corresponding buttons. A computer and the eye-tracking software record participants' eye movements during their reading, and their responses to the questions.

The main eye-tracking indexes in the Reading for Comprehension task vary according to different research questions. For those who are interested in readers' on-line sentence processing, important indexes are first fixation duration, gaze time, total reading time, and regression. First fixation duration is the duration of the first fixation readers land in a given region, which reflects the very early stage of word processing. Fixation durations are influenced by a number of low-level (visual) and high-level (cognitive or linguistic) factors, such as the length, frequency and predictability of the currently fixated word (Kliegl, Nuthmann, & Engbert, 2006). Longer first fixation duration always relates to less expected words, or more complex or less familiar words (Frenck-Mestre, 2005; Rayner, 1998). Furthermore, longer fixations in an ungrammatical condition also relate to an increased processing load, which is regarded as evidence of readers' sensitivity to the violation (Keating, 2009; Lim & Christianson, 2014). Gaze duration or gaze time is the total time of all fixations in a given region, from when readers' fixation first lands in this region until their eyes exit to the right or left. Usually, gaze duration is correlated with the first fixation duration, and it also reflects the early stage of processing. Gaze duration increases as the processing becomes more difficult (Frenck-Mestre, 2005; Rayner, 1998). Total reading time is the sum of all the fixations landed in a given region, reflecting participants' late stage of on-line processing and understanding of certain words (Frenck-Mestre, 2005; Rayner, 1998). Regression happens when readers land their fixation back to a given region that they have already passed. Regression reflects readers' late stage of processing and normally happens when readers encounter some unexpected items, and need to go back to get cues or more information to help them disambiguate certain parts of sentences (Frenck-Mestre, 2005; Rayner, 1998).

Among several eye-tracking reading studies which examined on-line sentences processing, garden-path sentences (e.g. 'While the man hunted the deer ran into the woods') have been most widely investigated. Normally when participants parse this sentence, they read from left to the right. When they reach 'the deer', they treat it as the object of 'hunted' as they have a preference for resolving the internal argument of the verb as soon as possible, and 'the deer' is the first possible object of 'hunted'. Thus the syntactic structure with 'the deer' as the object of 'hunted' has been built under this

analysis. However, when they keep reading to ‘ran into the woods’, they find out that the phrase ‘ran into the woods’ needs a subject, and ‘the deer’ is the only possible subject for this phrase in this sentence. At this point, readers start to realize that the syntax structure they have built before may not be correct, and need to read back to find more cues, or pause here to re-analyse the whole structure of the sentence. This whole process of building an initial structure - finding problems - reanalysing - fixing and building a new structure can be reflected in eye-tracking measure indexes. Participants present long fixation durations at ‘ran’ (the item that let participants realize their original analysis was wrong) and the following three words (spill-over effect), and read backward to the first part of the sentence (regressions) to look for more clues to reanalysis and build a correct syntactic structure (Christianson, Hollingworth, Halliwell, & Ferreira, 2001; Clifton, Staub, & Rayner, 2007). All these unconscious eye movement behaviours (e.g. longer fixation durations, word skip, regressions) reflect readers’ on-line processing of syntactic structures, which are implicit and automatic to native speakers.

Compared to alphabetic languages (e.g. English), less is known about Chinese speakers’ eye movement during on-line processing of Chinese sentences. As a logographic writing system, it is highly possible that Chinese readers have a very different eye movement pattern compared to English (Rayner, Li, Williams, Cave, and Well, 2007). It is important to use the eye-tracking Reading for Comprehension task to look into how people on-line process Mandarin Chinese.

In order to investigate how contextual information affects people’s understanding of ambiguous phrases, Zhang, Shu, Zhang, and Zhou (2002) used the Reading for Comprehension task to look into Chinese speakers’ understanding of ambiguous phrase – ‘VP+N1+*de*+N2’. ‘VP+N1+*de*+N2’ is a typical ambiguous phrase in Mandarin, it can either be interpreted as a Modifier-Noun Construction (MNC), or a Narrative-Object Structure (NOS). Some of these phrases are balanced between MNC and NOS, but others are either preferred to be interpreted as MNC (MNC-biased phrase) or NOS (NOS-biased phrase). Zhang et al. (2002) used both balanced phrases and NOS-biased phrases to look into the influence of the referential discourse context on participants’ on-line processing of these ambiguous phrases. They used a self-paced reading paradigm in Experiment 1 and an eye-tracking Reading for Comprehension task in Experiment 2. Participants were asked to read short passages for comprehension. Their behaviours and eye movements were recorded by a computer. The results showed that in both Experiment 1 and



Experiment 2 the referential discourse context affected participants' on-line processing of ambiguous phrases even at the early stage of processing: the two-referential context led to shorter reaction times and fixation durations on ambiguous phrases compared to the one-referential context. These results indicated that in Mandarin reading, previous context affects readers' on-line processing of ambiguous phrases. On processing Mandarin sentences, the eye-tracking Reading for Comprehension task can be used to collect data which can reflect readers' early and late stages of processing, and their ability of taking contextual information to make predictions in real time.

Wang, Pomplun, Chen, Ko, and Rayner (2010) used an eye-tracking Reading for Comprehension tasks to investigate how word predictability and other properties (number of strokes, frequency, etc.) affect native Mandarin speakers' on-line processing of contexts. Participants were asked to read a passage for comprehension, and answer two related questions about the content of the passage after finishing reading. A computer recorded their eye movements during the reading procedure. The manipulated variables of the target words are: the total number of strokes, word length, the average number of strokes, word frequency, and predictability. They found that the word lexical frequency effect was significant for first fixation duration and gaze duration, and was marginally significant for total reading time, which demonstrates that words' lexical frequency affects both the early and the late stages of on-line processing. Different from English (Rayner, 1998), first fixation durations are not affected by word length in Chinese, but are influenced by the average number of strokes: the more strokes the word has, the longer the fixation duration is. This observation indicates that the number of strokes of a word affects the early stage of on-line processing. As for the word predictability, there is a significant effect on first fixation durations, and a marginally significant effect on gaze durations, which suggests that word predictability can affect the early stages of on-line processing in Chinese reading.

The main goal of the current research is to explore how different nominal phrase structures (i.e. the mass/count syntactic cues) affect native Mandarin speakers' and late Dutch-Mandarin learners' on-line processing and interpretation, thus time-sensitive and space-sensitive information about participants' on-line processing is needed. In the Reading for Comprehension task, first fixation duration could help us gain a clear picture of the very early stage processing of each item in the nominal phrase (automatic and implicit processing), and the predictability of a certain classifier or a certain noun in a

specific nominal phrase structure. Total reading time could offer us information about the late stage of on-line processing and how difficult the procedure of understanding a certain region is. In addition, the Reading for Comprehension task is asking participants to read the sentences for comprehension. Participants are required to focus on the meanings of the sentences instead of the grammar, which could help us to look into participants' unconscious processing of nominal grammar and mass/count syntax, excluding all the potential influences caused by explicit strategy. In general, the Reading for Comprehension task based on the eye-tracking techniques is ideal for the current study. The detailed information about the eye-tracking Reading for Comprehension experiment in the current research is reported in Chapter 5.

### **2.3. The workload difference**

To investigate whether high proficiency late L2ers' on-line processing performance is limited by their cognitive resources (Performance Deficit Account vs. Representation Deficit Account), two experimental paradigms with different workload requirements were used in the current research. The difference of workload requirements between the two experiments is caused by three factors: the input modality, the task request, and the salience of the materials, which I outline in the following subsections.

#### ***2.3.1. The input modality***

First of all, it is the input modality which causes the processing load difference between the Visual World Paradigm and the Reading for Comprehension task. In the Visual World Paradigm, the materials were presented through two modalities: the visual images and the audio sentences. In the Reading for Comprehension task, on the other hand, the materials were presented through one modality: the visual written texts.

Paivio (1990) and Baddeley (1992) suggested that the working memory contains two separate processing channels for auditory and visual information. The auditory system is responsible for processing verbal information while the visual system is dealing with visual images and written texts. These two processing systems are independent and the capacity of each of them is limited. Since these two systems are independent and separate, the amount of information that can be processed by working memory can be enlarged if input information is presented in a mixed mode (auditory and visual) rather than in a single mode (auditory only, or visual only). In other words, the effective size of working memory can be increased by presenting materials in a multi-modality way (Allport,

Antonis, & Reynolds, 1972; Frick, 1984; Mousavi, Lou, & Sweller, 1995; Penney, 1989). Mayer (2005, 2009) further proposed the Cognitive theory of multimedia learning, which states that when presenting pictures and text together, the text that is presented auditorily will lead to better learning and understanding results than when the text is presented visually. This is because the information of both the pictures and the written texts is processed through the visual system, while the information of audio material texts is processed through the auditory system. Since both the visual and the auditory systems are of limited capacity and can only process a certain amount of information at a time, to use both systems (visual pictures and audio texts) instead of overloading a single one (visual pictures and visual written texts) is advantageous.

Previous research found that the multi-modality effect has been spotted on children learning and comprehension (Ginns, 2005; Herrlinger, Höffler, Opfermann & Leutner, 2016; Levie & Lentz, 1982; Levin et al., 1987; Peeck, 1994; Segers et al., 2008; Tabbers et al., 2004), and adult learning and processing (Harskamp et al., 2007; Mayer, 2009; Mayer & Fiorella, 2014; Sweller et al., 2011; Tabbers et al., 2004; Tindall-Ford et al., 1997). The facilitative effect of the multi-modality in L2 learning and comprehension has also been found in a lot of studies. Baltova (1999) found that the addition of images to audio materials helps L2ers to set the scene of events which leads to their better learning results. Guichon & McLornan (2008) investigated the effects of multi-modality on L2 comprehension. They manipulated the information input to have four conditions: audio materials only; audio materials with images; images, audio materials and L1 subtitles; and images, audio materials, and L2 subtitles. Participants were asked to produce a detailed written summary in English based on their understanding of the input materials and with the help of their own notes. They found that participants' comprehension improved when they were exposed to a text in several modalities. The modality effect can benefit L2ers' comprehension. Mayer, Lee, & Peebles (2014) found that inputs from visual pictures and audio materials can be better learned and processed by L2ers than inputs from only visual written texts, which has been confirmed by numerous previous studies (Gu, 2013; Hinkel, 2013; Min, 2008; Plass & Jones, 2005; Plass, Chun, Mayer, & Leutner, 1998; Wang, 2011).

In the current study, the Visual World Paradigm experiment presents materials through a multi-modality way (visual pictures and audio materials) while the Reading for Comprehension experiment presents materials through a single visual modality (i.e.

written texts). Thus, the amount of the information can be processed by working memory in the Visual World Paradigm experiment is larger than in the Reading for Comprehension experiment, which then leads to higher processing load in the latter than in the former on participants.

### *2.3.2. The task request*

The Visual World Paradigm experiment also differs from the Reading for Comprehension experiment in the task request. In the Visual World Paradigm, participants are asked to listen to the instructions from the headphones, and choose one out of four pictures on the screen based on the instructions they heard. During the experiment, participants need to pay attention to the critical nominal phrases and choose a target picture based on the audio input of every item of the nominal phrases. On the other hand, in the Reading for Comprehension experiment, participants are required to read the sentences silently for comprehension, and answer a meaning-related question after each sentence. During the experiment, participants need to pay attention to the critical nominal phrases as well as the whole sentences, and be prepared for the upcoming questions. The procedure of the picture-selection based on the audio input in the Visual World Paradigm experiment is easier and less processing load demanded than the procedure of the reading and understanding sentences in the Reading for Comprehension experiment. In other words, the Visual World Paradigm experiment requires a lower workload than the Reading for Comprehension experiment.

### *2.3.3. The salience of the key materials*

Apart from the material input modality and the task request, the two experiments also differ from each other in the salience of the key materials. In the Visual World Paradigm experiment, the audio instruction has the structure ‘from the four pictures, please choose X’, in which X refers to the critical nominal phrase<sup>8</sup>. While in the Reading for Comprehension experiment, the written sentence has the structure ‘there is a X in A, A is ...’, in which X is the critical nominal phrase, while A is an appropriate location phrase. The question after each sentence always refers to the location phrase A instead of the

---

<sup>8</sup> Detailed information about the materials in the Visual World Paradigm experiment is presented in Chapter 4.

critical nominal phrase X<sup>9</sup>. The critical nominal phrases occur at the end of the audio input in the Visual World Paradigm experiment, but in the middle of the written sentence in the Reading for Comprehension experiment. Thus it is easier for participants to keep the information of the critical nominal phrase in the working memory system for a relatively long time in the Visual World Paradigm experiment than in the Reading for Comprehension experiment (the Retroactive Interference effect, Müller, 1900; Underwood, 1948)<sup>10</sup>. Furthermore, in the Visual World Paradigm experiment, the audio instruction and the four pictures occurred at the same time in each trial. The four pictures in each trial can be treated as visualized possible interpretations of the critical nominal phrase, which makes it salient and easy for participants to understand the critical nominal phrase. However, in the Reading for Comprehension experiment, the key nominal phrases and the sentences were presented as written texts. It is obvious that compared to written texts, pictures and images are more attention-attractive and straightforward. Thus the key materials in the Visual World Paradigm experiment are more salient and explicit for participants than in the Reading for Comprehension experiment. And consequently, the Visual World Paradigm experiment requires less processing load than the Reading for Comprehension experiment.

Based on above discussion we can see that, from the three aspects (the input modality, the task request, and the salience of the key materials), the Visual World Paradigm and the Reading for Comprehension task require different workloads on participants. Compared to the Reading for Comprehension experiment, the Visual World Paradigm experiment requires a lower workload on participants. By comparing high proficiency late Dutch-Mandarin learners' real-time processing performance between these two experiments, one can tell whether their performance is limited by their cognitive resources. According to the Fundamental Difference Approach, especially the Representation Deficit Account, late L2ers' processing difficulties are caused by their

---

<sup>9</sup> Detailed information about the materials in the Reading for comprehension experiment is presented in Chapter 5.

<sup>10</sup> Retroactive interference occurs when newly learned information interferes with and impedes the memory and recall of previously learned information, resulting in a decreased ability of recalling primary learned information due to the latest learning of new information. The Retroactive Interference effect predicts that it is easier and less-effort needed for people to remember the newly-learned/ latter-occurred knowledge than the previously-learned/ prior-occurred knowledge.

deficient representations. The change of workload requirements would not affect L2ers' on-line processing. Thus in the current research, the Representation Deficit Account predicts that high proficiency late Dutch-Mandarin learners will exhibit similar behaviour patterns between the Visual World Paradigm experiment and the Reading for Comprehension experiment. On the other hand, the Fundamental Similarity Approach, especially the Performance Deficit Account claims that late L2ers' processing is limited by their cognitive resources and L2 proficiency. With high L2 proficiency and a low workload requirement, they are able to exhibit native-like behaviours. Thus the Performance Deficit Account predicts that in the current research, it is more likely for high proficiency late Dutch-Mandarin learners to exhibit native-like behaviours in the Visual World Paradigm experiment (with a comparatively lower workload) than in the Reading for Comprehension experiment (with a comparatively higher workload).

To summarise, in the current research, the Fundamental Difference Approach and the Fundamental Similarity Approach make different predictions on how high proficiency late Dutch-Mandarin learners will behave in the two experiments. The results of the current research offer some evidence to support the Fundamental Similarity Approach that late Dutch-Mandarin learners exhibited native-like behaviours in the Visual World Paradigm experiment, but not in the Reading for Comprehension experiment. The detailed information of the Visual World Paradigm experiment and the Reading for Comprehension experiment is introduced in Chapter 4 and Chapter 5 respectively, and the general discussion on the results of the two experiments is presented in Chapter 6.

## **IV. Visual Word Paradigm Experiment**

In this chapter, detailed information about the Visual World Paradigm experiment is presented. As mentioned in Chapter 1, this current research is aiming to look into how native Mandarin speakers and high proficiency late Dutch-Mandarin learners on-line use mass/count syntactic cues to build anticipations. As discussed in Chapter 2, the classifier system and the mass/count syntactic cues are unique-to-L2 constructions for Dutch-Mandarin learners. By comparing their behaviours to native Mandarin speakers', one can tell whether high proficiency L2 learners could acquire unique-to-L2 constructions through implicit learning, and whether they could have native-like behaviours during on-line processing and interpretation. Based on the literature review presented in Chapter 3, most of the previous studies which looked into the processing of classifiers and the mass/count distinctions focused on native Mandarin speakers' use of different classifiers by using some off-line tests (e.g. Chen et al., 2003; Gong, 2010; Liang, 2009). Very few of them investigated L2ers' acquisition of classifiers and the mass/count distinction in Mandarin (Lau & Gruter, 2015). So far Li, Barner & Huang (2008) is the only one experimental study which used off-line test to investigate the influence of mass/count syntax on native Mandarin speakers. In spite of Li et al. (2008)'s contributions to the field of mass/count syntax processing in Mandarin, there are some improvements which can be done to fetch a clear picture of native Mandarin speakers' and L2-Mandarin-learners' on-line processing of mass/count syntax. As mentioned in Chapter 3, the Visual World Paradigm experiment is conducted to fill these gaps.

This chapter is structured as follows. In Section 1, materials' norming and manipulation are presented. In Section 2, participants' information is reported. In Section 3, the procedure of the Visual World Paradigm is introduced. In Section 4, predictions are raised. In Section 5 detailed results are reported. In Section 6, discussions are laid out.

### **1. Material**

Following Li et al. (2008), the classifiers *gen* ‘rod’, *kuai* ‘chunk’, and *ba* ‘handful’ were used in this Visual World Paradigm experiment. Considering the fact that materials used in the Visual World Paradigm need to be visualized in appropriate pictures, the classifier *pian* ‘slice’ was excluded in the current experiment because some nouns occurring with it cannot be properly picturized<sup>11</sup>. As mentioned in Chapter 2, all of the three classifiers are dual-role classifiers since they have two possible interpretations: the count-classifier reading and the massifier reading.

According to Cheng & Sybesma (1998), there are two syntactic cues for massifiers: the classifier’s ability to be modified by adjectives (Adj), and the insertion of the modification marker *de* after the classifier. The current experiment focused on the influence of the Adj-CL word order on participants’ on-line processing. There are three conditions of nominal phrase structures: C1 (Condition 1) with the structure [Num+CL+N] as the baseline, C2 with the structure [Num+CL+Adj+N], and C3 with the structure [Num+Adj+CL+N]. Comparing participants’ reactions to nominal phrases in C1 and C2, one can tell the influence of adding an Adj on real time processing, while the effect of the word orders (CL-Adj vs. Adj-CL) can be revealed by comparing participants’ reactions between C2 and C3.

Nominal phrases (NPs) were embedded in sentences with the structure ‘From the four pictures, could you please choose + [NP]’. By comparing participants’ fixations distributions on the four pictures among the three conditions of nominal phrases, it can be revealed that how different nominal phrase structures affect native Mandarin speakers’ and Dutch-Mandarin learners’ on-line interpretation and anticipation building. It should be noted that apart from the cardinal meaning, the numeral *yi* ‘one’ in Mandarin can also have a ‘whole’ meaning (Chao, 1968; Cheng & Sybesma, 1998; Paris, 1981). In order to eliminate this Generalizing effect the numeral *yi* ‘one’ may cause, in the current experiment, the numeral *san* ‘three’ was used in each nominal phrase. In addition, the adjectives *da* ‘big’ and *xiao* ‘small’ were used in nominal phrases. In each list, half of the nominal phrases were with the adjective *da* ‘big’, and the other half were with the adjective *xiao* ‘small’.

---

<sup>11</sup> The classifier *pian* ‘slice’ was included in the Reading for Comprehension experiment reported in Chapter 5.



Considering one of the main concerns of the current research is investigating whether native Mandarin speakers and late Dutch-Mandarin learners can take advantage of mass/count syntactic cues to make predictions in real time, the mass/count status of the nouns was manipulated. In order to select typical count nouns as well as mass nouns, a Noun Rating Test was conducted before the Visual World Paradigm experiment on native Mandarin speakers<sup>12</sup>.

### 1.1. Noun Rating Test: The mass/count categorization of nouns

Among a word pool containing around 90 high frequency simple nouns which can normally occur with one of the four dual-role classifiers used in Li et al. (2008) (*gen* ‘rod’, *kuai* ‘chunk’, *pian* ‘slice’, and *ba* ‘handful’), ten native Mandarin speakers (who did not participate in the Visual World Paradigm experiment) were required to judge whether a noun is a typical count noun or a typical mass noun on a 5-points rating scale using the mass noun feature ‘dividability’ from Krifka (1989), in which 1 stands for ‘dividable’ while 5 stands for ‘undividable’. If the entity denoted by a noun can be divided several times, and each part of it after being divided still has the property the original entity has, the noun is ‘dividable’. On the contrary, if the entity denoted by a noun cannot be divided, or each part of it after being divided possesses different features from the original entity, the noun is ‘undividable’. Based on the results of the rating, 14 nouns were rated as typical mass nouns (mean rating score = 1.58), and 28 nouns were rated as typical count nouns (mean rating score = 4.39). All the chosen typical count nouns and mass nouns are presented in Appendix A.1. Among these nouns, 12 typical mass nouns and 12 typical count nouns which can occur with the one of the three dual-role classifiers used in the Visual World Paradigm experiment were chosen. Each noun contains two characters; the number of syllables of each character and the lexical frequency of each noun were controlled. Eight nouns occurred with the classifier *ba* ‘handful’ (four count nouns & four mass nouns), ten nouns occurred with the classifier

---

<sup>12</sup> This Noun Rating Test was conducted only on native Mandarin speakers, but not on Dutch-Mandarin learners. Based on the discussion in Chapter 2 that speakers of different languages share the same count/mass perceptions of objects/substances in the world (Barner, Inagaki & Li, 2009; Choi, Ionin & Zhu, 2017; Gleitman & Papafragou, 2005; Li, Dunham & Carey, 2009; Li & Gleitman, 2002; Mazuka & Friedman, 2000). In the current study, I argue that speakers of Mandarin and Dutch should have the same count/mass mental representations of objects/substances. Thus the typical count/mass nouns normed by native Mandarin speakers, should also be typical count/mass nouns to Dutch speakers. Further count/mass noun rating test should be conducted on Dutch-Mandarin learners to confirm this argument.

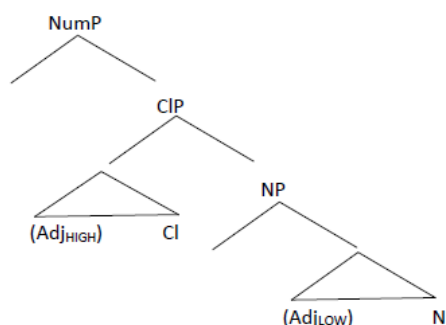
*kuai* ‘chunk’ (five count nouns & five mass nouns), and six nouns occurred with the classifier *gen* ‘rod’ (three count nouns & three mass nouns). For each pair of nouns (a count and a mass noun which share the same classifier), in order to avoid the Phonological Competition Effect (Klein, Carlson, Li, Jaeger, & Tanenhaus, 2012) and Tone Sandhi Effect (Yip, 2002), the first character of each noun was controlled to have the same tone (and most of them share the same first syllable).

For each pair of nouns, there were four pictures presented on the screen. Two of them contained entities denoted by mass nouns, and the other two contained objects denoted by count nouns. In addition, based on the semantic difference between the classifier *ba* and the classifiers *gen* and *kuai* (*ba* is a collective classifier when being a massifier, while *gen* and *kuai* are dividers when being massifiers, also see Cheng, 2010), entities on the pictures were organized in different ways. In trials with the classifier *ba* ‘handful’, one of the two count-noun denoting pictures contained three solid individual objects (e.g. three spoons), the other one contained three groups of the same objects (e.g. three handfuls of spoons). The former picture is named as ‘count-count picture’ since it is a count-noun denoting picture in which the objects are presented in solid and individual units, while the latter one is named as ‘count-mass picture’ since it is a count-noun denoting picture in which objects are organized in plural sets. The two mass-noun denoting pictures both contained three groups of entities. The difference between these two pictures is either the size of the group is consistent with the adjective in the nominal phrase (e.g. three big/small handfuls of pebbles), or the size of the individual entity is consistent with the adjective but the size of the group is not (e.g. three handfuls of big/small pebbles). This design is corresponding to the nominal phrases with the structure [Num+CL+Adj+N] and [Num+Adj+CL+N]. To be specific, in Condition 2 with the structure [Num+CL+Adj+N], the adjective is preceding and modifying the noun, while in Condition 3 with the structure [Num+Adj+CL+N], the adjective is preceding and modifying the classifier. The mass-noun denoting picture in which the size of the group is consistent with the adjective (e.g. three big/small handfuls of pebbles) is named as ‘mass-high-Adj’ picture since it is referring to the nominal phrase in which the adjective is attached to and modifying the classifier (the nominal phrase in Condition 3). On the other hand, the mass-noun denoting picture in which the size of the individual entity is consistent with the adjective but not the size of the group (e.g. three handfuls of

big/small pebbles) is named as ‘mass-low-Adj’ picture since it is referring to the nominal phrase in which the adjective is attached to and modifying the noun (the nominal phrase in Condition 2). The use of the terms ‘high-Adj’ and ‘low-Adj’ is based on the positions of the adjectives when they modifying different items: compared to modifying the nouns in Condition 2, the adjective occurs at a higher place in the syntactic structure when modifying the classifiers in Condition 3<sup>13</sup>. Thus in each trial, there were four pictures presented on the screen at the same time: the count-count picture (the count-noun denoting picture with solid objects organized in their natural units), the count-mass picture (the count-noun denoting picture with solid objects organized in groups), the mass-high-Adj picture (the mass-noun denoting picture with groups of substances and the size of the group is consistent with the adjective), and the mass-low-Adj picture (the mass-noun denoting picture with groups of substances and the size of the individual unit is consistent with the adjective).

In trials with the classifier *kuai* ‘chunk’ and *gen* ‘rod’, one of the two count-noun denoting pictures contained three individual objects (e.g. three units of discrete shaped jade), the other one contained three divided parts of the same objects (e.g. three chunks of shaped jade). Thus the count-count picture is the count-noun denoting picture in which objects were presented in individual units, the count-mass picture is the count-noun denoting picture in which objects were presented in divided mass portions. The two mass-noun denoting pictures both contained three portions/units of substances. The difference between these two pictures is either the size of the portion is consistent with the adjective in the nominal phrase (e.g. three big erasers), or the size of the portion is inconsistent with the adjective (e.g. three small erasers). Thus in each trial, there were four pictures presented on the screen at the same time: the count-count picture (the count-

<sup>13</sup> This is because in the syntactic structure of a nominal phrase, the classifier projection (CIP) is always higher than the noun projection (NP), as illustrated below.



noun denoting picture with objects which are organized in their natural units), the count-mass picture (the count-noun denoting picture with objects organized in divided parts), the mass-high-Adj picture (the mass-noun denoting picture with substances organized in Adj-consistent-sizes), and the mass-low-Adj picture (the mass-noun denoting picture with substances organized in Adj-inconsistent-sizes). The examples of the material are illustrated in (34).

(34). a. Trials with the classifier *ba* ‘handful’

C1: *san* *ba* *shaozi/ shizi*

three **CL** spoon/ pebble

‘three spoons/ three handfuls of spoons/pebbles’

C2: *san* *ba* *da* *shaozi/ shizi*

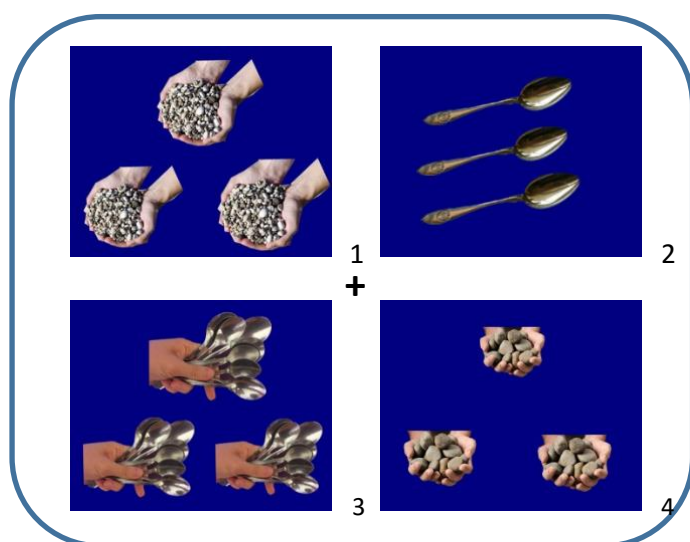
three **CL** **big** spoon/ pebble

‘three big spoons/ three handfuls of big spoons/pebbles’

C3: *san* *da* *ba* *shaozi/ shizi*

three **big** **CL** spoon/ pebble

‘three big handfuls of spoons/ three big handfuls of pebbles’



1. mass-high-Adj
2. count-count
3. count-mass
4. mass-low-Adj

b. Trials with the classifier *kuai* ‘chunk’ (‘Units’ in these trials stands for either the individual units objects naturally possess, or created measure units which can be described as ‘chunk’.)

C1: *san* *kuai* *yupei/ dangao*

three **CL** jade/ cake

‘three units of shaped jade/ three units of cake’

C2: *san* *kuai* *da* *yupei/ dangao*

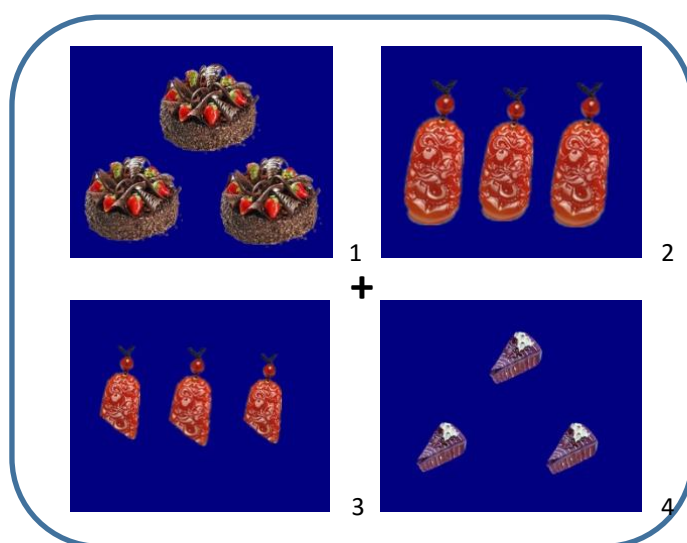
three **CL** **big** jade/ cake

‘three units of big shaped jade/ three units of big cake’

C3: *san* *da* *kuai* *yupei/ dangao*

three **big** **CL** jade/ cake

‘three big units of shaped jade/ three big units of cake’



1. mass-high-Adj
2. count-count
3. count-mass
4. mass-low-Adj

c. Trials with the classifier *gen* ‘rod’ (‘Units’ in these trials stands for either the individual units objects naturally possess, or created measure units which can be described as ‘long and thin’.)

C1: *san* *gen* *huanggua/ shengzi*

three **CL** cucumber/ string

‘three units of cucumbers/ strings’

C2: *san* *gen* *xiao* *huanggua/ shengzi*

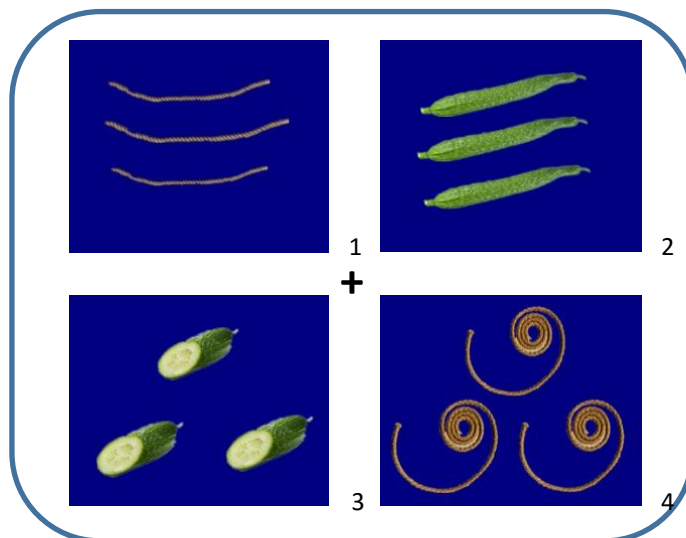
three **CL** **small** cucumber/ string

‘three units of small cucumbers/ three units of small strings’

C3: *san* *xiao* *gen* *huanggua/ shengzi*

three **small** **CL** cucumber/ string

‘three small units of cucumbers/ three small units of strings’



1. mass-high-Adj
2. count-count
3. count-mass
4. mass-low-Adj

Overall, there were 72 sentences in total (24 nouns \* 3 conditions). All the sentences were divided into three lists pseudo-randomly to make sure that each list only contained one of the three conditions of each sentence. There were 24 fillers in each list which makes the total number of sentences in each list is 48. These 144 sentences (72 critical audio materials + 72 fillers) were digitally recorded by a female native speaker of Mandarin in a sound-proofed booth, sampling at 44.1 KHz. In each of the critical audio sentences, the normed durations for the numeral is 300ms, for the classifier is 250ms, for the adjective is 300ms, and for the noun is 350ms. Each participant only tested on one of the three lists. The order of the four pictures in each trial were counterbalanced in each list. The whole experiment lasted around 30 minutes.

In order to make sure that all the pictures in the Visual World Paradigm experiment are recognizable and understandable to participants, a Norming Test on native Mandarin speakers and a Naming Test on Dutch-Mandarin learners were conducted before the Visual World Paradigm experiment.

### **1.2. Norming Test**

In the Norming Test, ten native Mandarin speakers (who did not participate in the Visual World Paradigm experiment and the Noun Rating Test) were asked to judge whether the pictures are recognizable to them or not by using a 5-point scale, in which 1 stands for unrecognizable (not be able to tell what is the entity on the picture), while 5 stands for recognizable (can easily tell what is the entity in the picture). The results showed that the mean scores for all the pictures were above 4.5, which means all the pictures are recognizable and acceptable to native Mandarin speakers.

### **1.3. Naming Test**

To make sure that all the pictures are recognizable and understandable to L2 participants in the current study, all the high proficiency late Dutch-Mandarin learners (the L2 participants who participated in the Visual World Paradigm experiment and the Reading for Comprehension experiment) were asked to name all the pictures by writing down both English and Mandarin names, and speak out loud the names of the objects in Mandarin. The accuracy of their writing answers and their pronunciations of the pictures were recorded and analysed. The results showed that all the pictures were recognizable and understandable to L2 participants. They were familiar with the objects and the substances presented in all the pictures, and were able to write down and speak out the

right names (they were familiar with both the writing characters and the pronunciations of the Chinese names). Through this Naming Test, it is assured that the high proficiency late Dutch-Mandarin learners in the current study are familiar with all the nouns (both the written names and pronunciations in Chinese), and that all the pictures are recognizable and acceptable to them.

## 2. Participants

Two groups of participants participated in the Visual World Paradigm experiment: native Mandarin speakers (L1ers) and high proficiency late Dutch-Mandarin learners (L2ers). There were 30 native Mandarin speakers (15 females, 15 males). They were all students from Beijing Normal University, aged between 18 to 28 years old. None of them had participated in the Noun Rating Test or the Norming Test. Each participant was given ¥ 50 for their participation. The data of the native Mandarin speaker group was collected in Beijing Normal University, Beijing, China.

The L2 learner group consisted of 30 Dutch-Mandarin learners (16 females and 14 males). They were students from Leiden University (aged between 16 to 35 years old). They all started learning Mandarin after puberty. All of them had been learning Mandarin for at least 3 years and had an experience of studying in China for at least 6 months. All L2 participants had passed the HSK-C (Hanyu Shuiping Kaoshi – advanced level, the standard Chinese language proficiency test for non-native speakers administered by the Ministry of Education of the People's Republic of China), indicating that they are high-proficiency L2-Mandarin learners. All L2 participants were asked to complete a 5-point-scale self-rating questionnaire to report their ability of Mandarin listening, speaking, reading, and writing, in which 1 stands for poorly used, while 5 stands for very fluent. The average scores for the four language abilities were 4.2, 3.0, 4.5, and 3.1 respectively, which indicated that these Dutch-Mandarin learners see themselves as fluent Mandarin users in listening and reading, but intermediate users of Mandarin in speaking and writing. Each L2 participant was given €15 for their participation. The data of the L2 learner group was collected in Leiden University, Leiden, Netherland.



### 2.1. Blank-filling Test

From the Naming Test we can see that the high proficiency late Dutch-Mandarin learners are familiar with all the nouns (both the written names and the pronunciations) used in the current study. To further make sure that all the L2 participants are familiar with the classifier-noun pairs used in the current study, a Blank-filling test was conducted on the same group of L2ers. In the Blank-filling Test, each noun (the typical count and mass nouns chosen in the Noun Rating Test) was presented with the numeral *yi* ‘one’ and a pair of brackets preceding it, in a structure like [one (    ) N]. Dutch-Mandarin learners were asked to fill in appropriate classifiers in the brackets. They could write down multiple classifiers if they want to. The results showed that most of the first classifier L2 participants filled in for each noun is the classifier used in the current study. Even if for some nouns, the first filled-in classifier is not the classifier used in current study, the classifiers listed in the second or third place are. This result indicated that the Dutch-Mandarin learners in the current research are familiar with the classifier-noun pairings used in the current research. Two out of 30 Dutch-Mandarin participants used the general classifier *ge* for more than a quarter of the nouns. Based on previous studies (Polio, 1994), this over-generalized use of classifier *ge* indicated that these two Dutch-Mandarin learners had not fully acquired the classifier-noun pairings. Hence, these two L2 participants’ data were excluded from the final analysis. This Blank-filling Test reveals that high proficiency Dutch-Mandarin learners have already acquired the basic and necessary lexical knowledge of the classifiers (the classifier-noun pairs) in the current research.

In the current research, the Dutch-Mandarin learners participated in the Visual World Paradigm experiment also participated in the Reading for Comprehension experiment, the Naming Test, and the Blank-filling Test. This is because the number of the available high proficiency late Dutch-Mandarin learners at that time was limited. This may be considered problematic since it may cause some priming/learning effect due to the repeated appearance of similar materials. In order to reduce the possible priming/learning effect on L2 participants, the order of the two eye-tracking experiments and off-line tests was manipulated to be counterbalanced: the Naming Test → the Reading for Comprehension experiment/the Visual World Paradigm experiment → the Blank-filling Test → the Visual World Paradigm experiment/the Reading for Comprehension experiment. Half of the L2 participants tested on the Visual World Paradigm experiment

before they participated in the Reading for Comprehension experiment, while the other half participated in the Reading for Comprehension experiment first. There was a 10-minutes break after each task.

### 3. Procedure

SR Research Eyelink 1000 was used to measure participants' eye movements. Both left and right eyes' movements were recorded, but only the right eye's data were analysed. Following the nine-points calibration and validation, gaze-position error was less than 0.5°. Participants were tested in a sound-proof booth and seated 60cm from a 19-inch monitor.

Participants were tested individually. Before the critical experimental trials, there were instructions and 10 practice trials. After participants read the instruction and finished the practice trials, a standard 9-point grid calibration and validation was completed. During the experiment, participants were asked to listen to the sentences from the headphones. At the same time, in each trial, there were four pictures presented on the screen. Participants were required to choose one of the four pictures based on the sentences they heard from the headphones by moving the mouse to click the corresponding picture. Participants' gaze was directed to a fixation cross in the middle of the screen prior to each trial to avoid baseline effects of participants already looking at a certain item before the onset of the audio inputs (Barr et al., 2011; Hopp, 2016). A trial only started when participants fixed on the calibration dot stably. Participants' eye movements during the display of the audio materials and their responses were recorded by the eye-tracking software and the computer. The ethical protocol approval was obtained from Beijing Normal University and Queen Mary Ethics of Research Committee.

### 4. Prediction

According to the discussions in Chapter 2 that in Mandarin Chinese, count-classifiers often occur with count nouns while massifiers can occur with either count nouns or mass nouns. In addition, based on the findings in Li et al. (2008) that native Mandarin speakers

chose either non-solid substances or multiple solids when hearing nominal phrases in the structure [Num+Adj+CL+*de*+N], while choosing individual solids on hearing the structure [Num+CL+Adj+N]. In this current research I predict that with dual-role classifiers, nominal phrases with different mass/count structures (mass/count syntactic cues) will have different interpretations. To be specific, the mass syntactic cue (i.e. the Adj-CL word order) will force a mass/plural meaning on the nominal phrase for native Mandarin speakers.

In Condition 1 where the nominal phrases have the structure [Num+CL+N], native Mandarin speakers are expected to land their fixations randomly around the four pictures until the onset of the noun. Since there is no mass/count syntactic cue in the nominal phrase, and the classifier is a dual-role classifier which can be either a count-classifier or a massifier, the only informative cue in the the structure [Num+CL+N] is the noun. Participants will have to wait till the onset of the noun to decide which is the target picture and then direct their fixations to it. When they heard count nouns, they should focus on the count-noun denoting pictures (the count-count picture and the count-mass picture), and choose one out of them based on their interpretations of the nominal phrases. Similarly, when they heard mass nouns, they should focus on the mass-noun denoting pictures (the mass-high-Adj picture and the mass-low-Adj picture), and choose one out of them based on their interpretations of the nominal phrases.

In Condition 2 where the nominal phrases have the structure [Num+CL+Adj+N], participants are expected to have randomly distributed fixations around the four pictures until the onset of the noun. Similar to the nominal phrases in Condition 1, in Condition 2 there is no mass/count syntactic cue offered (the structure [Num+CL+Adj+N] is a neutral structure without any preference for count or mass interpretations). The onset of the nouns should be the only cue for the target picture. When they heard count nouns, they should focus on the count-noun denoting pictures (the count-count picture and the count-mass picture) and choose the one in which the size of the entities is consistent with the adjectives they have heard. They should choose the count-count picture when the adjective is *da* ‘big’, while choosing the count-mass picture when the adjective is *xiao* ‘small’. This is because both of the units of the entities in the count-count picture (i.e. the solid individual unit) and the count-mass picture (i.e. the grouped/divided unit) can be described by the classifiers. The difference between these two pictures is the size of the unit. The entities in the count-count picture (i.e. the individual undivided objects)

always have bigger sizes than the ones in the count-mass picture (i.e. the grouped/divided objects). Thus participants should choose the count-count picture when the adjective is *da* ‘big’, while choosing the count-mass picture when the adjective is *xiao* ‘small’.

When they heard mass nouns, they should focus on the mass-noun denoting pictures (the mass-high-Adj picture and the mass-low-Adj picture) and choose the one in which the size of the individual entity is consistent with the adjectives. To be more specific, when the classifier is *ba*, they should choose the mass-low-Adj picture to be the target since the size of the individual entity in that picture is consistent with the adjectives. While when the classifier is *gen/kuai*, they should choose the mass-high-Adj picture since it is the only one mass-noun denoting picture in which the size of the entities is consistent with the adjectives.

In Condition 3 where the nominal phrases have the structure [Num+Adj+CL+N], the mass syntactic cue is offered: the Adj-CL word order. On hearing the Adj directly following the Num, if participants are sensitive to the mass syntactic cue in Mandarin, they should realize that the upcoming item is a massifier. At this point, they are expected to build an anticipation for the nominal phrase to have a mass or a plural interpretation. As a consequence, more fixations should be landed at the pictures which expressing mass or plural meanings (the mass-high-Adj picture, the mass-low-Adj picture, and the count-mass picture) on the onset of the Adj. In this case, the mass syntactic cue (i.e. the Adj-CL word order) is expected to be used by participants to narrow down the possible target options. The onset of the classifier should not change participants’ fixation distributions since it will only confirm that their previous massifier-anticipation built based on the onset of the adjectives is correct. On hearing count nouns, they should focus on the count-mass picture since this is the only picture which is compatible with both the massifier-anticipation and count-noun cue. On hearing mass nouns, they should focus on the mass-high-Adj picture and the mass-low-Adj picture, and finally choose the mass-high-Adj picture as the target since it satisfies the massifier-anticipation, the mass noun cue, and also the adjectives.

In general, the predictions of native Mandarin speakers’ fixation preference (the pictures with the highest proportion of fixations) among the four pictures along the input of nominal phrases are summarized in Table 7.

Table 7. Predictions of L1ers' fixation preferences upon hearing each element

	Num	CL	Count-N	Mass-N	
Condition 1	-	-	count-count or count-mass	mass-high-Adj or mass-low-Adj	
	Num	CL	Adj	Count-N	Mass-N
Condition 2	-	-	-	count-count with <i>da</i> ‘big’	mass-low-Adj with <i>ba</i>
				count-mass with <i>xiao</i> ‘small’	mass-high-Adj with <i>gen/kuai</i>
	Num	Adj	CL	Count-N	Mass-N
Condition 3	-	count-mass or mass-high-Adj or mass-low- Adj	count-mass or mass-high-Adj or mass-low- Adj	count-mass	mass-high-Adj

As for Dutch-Mandarin learners, the Feature Reassembly Hypothesis and the Morphological Congruency Hypothesis make different predictions. Based on the Feature Reassembly Hypothesis, in the procedure of 'feature reassembly', Dutch-Mandarin learners firstly need to notice that the mass/count distinction is marked at the classifier system, which is different from Dutch. After that, they need to learn that there are count-classifiers as well as massifiers, and they have different functions, meanings, and distributions. Further, they also need to know that there is a group of classifiers which could have either count-classifier meanings or massifier meanings. Embedding these dual-role classifiers in nominal phrases with different mass/count syntactic cues would lead to different interpretations. Considering the fact that the Dutch-Mandarin participants in the current research were high proficiency L2ers, they had not reached the end-state stage. It is possible that they may not fully acquire these dual-role classifiers' ambiguous meanings and the mass/count syntactic cues. The possibilities of what they

have already acquired and the corresponding expected behaviours are summarized in Table 8.

*Table 8. Predictions about L2ers' behaviours*

<b>What L2ers know</b>		<b>Their expected behaviours</b>
Only count-classifier meanings		Fixate the count-count picture in all three conditions on hearing count nouns
Only massifier meanings		Fixate the count-mass picture in all three conditions on hearing mass nouns
Dual-role classifiers	Only the lexical knowledge	Frequency/preference effect
	Both the lexical and the syntactic knowledge	Native-like patterns

To be specific, in the current study, on hearing count nouns, the Dutch-Mandarin learners should always look at and choose the count-count picture if they only know the count-classifier meanings of these dual-role classifiers, while always choose the count-mass picture if they only know the massifier meanings. In these two cases, the mass/count syntactic cue would not affect L2 participants' processing. If they already know that these are dual-role classifiers which could have either count-classifier or massifiers meanings, there are two possibilities. Firstly these Dutch-Mandarin learners may only acquire the lexical knowledge of these classifiers but not the syntactic knowledge. In this case, they are expected to display frequency/preference-consistent patterns. For example, if the classifier *ba* has a higher frequency of being interpreted as a count-classifier in daily input, while the classifier *gen* has a higher frequency of being interpreted as a massifier, then Dutch-Mandarin learners were expected to choose the count-count picture on hearing count nouns in all three conditions when the classifier was *ba*, and choose the count-mass picture on hearing count nouns in all three conditions when the classifier was *gen*. On the other hand, if Dutch-Mandarin learners have already acquired

both the lexical and the syntactic knowledge of classifiers, then native-like behavioural patterns are expected.

The Morphological Congruency Hypothesis, on the other hand, claims that even though L2ers have reached an advanced proficiency level of L2, they still cannot fully acquire the unique-to-L2 representations. To be specific, in the current research, high proficiency Dutch-Mandarin learners may not know that these dual-role classifiers could have either count-classifier meanings or massifier meanings. Or, they may not be sensitive to the mass/count syntactic cues during on-line processing. In either case, they would not be affected by nominal phrases with different structures. No difference will be expected among the three conditions.

In the current research, along with the Performance Deficit Account and the Feature Reassembly Hypothesis, I predict that high proficiency late Dutch-Mandarin learners can acquire unique-to-L2 constructions (i.e. the mass/count syntactic cues) through implicit learning. They can exhibit native-like behaviours in the predictive use of the mass/count syntactic cues.

## **5. Results**

There are two parts of results in the Visual World Paradigm experiment: participants' fixation distributions and their behavioural results. The proportions of fixations in each picture tell us how participants predictively process each item of the nominal phrases during real time, and the behaviour data reveals their final choices. Since the main concern of the current research is investigating whether native Mandarin speakers and Dutch-Mandarin learners can take use of mass/count syntactic cues during predictive processing, the key research question in the Visual World Paradigm experiment is how different nominal phrase structures affect participants' fixation distributions among the four types of pictures along the audio inputs. Thus the data of participants' fixation distributions is reported first, followed by participants' behaviour data.

### **5.1. Fixation distributions before the nouns**

Currently, there is no consensus on the best way to statistically test the observations in the Visual World Paradigm experiment. The difficulty with the Visual World Paradigm

data is that fixations in any given region and at any given time are categorical (i.e. 0 or 1), whereas the independent measure, i.e. the time sequence, is continuous. A variety of analyses have been used to examine differences in proportion fixations (see special issue 59 of the *Journal of Memory and Language*, 2008). Traditional models compare the proportional fixation to different objects/images over time using ANOVA or t-tests. The problem with such models is that they violate the underlying statistical assumptions of these tests (Barr, 2008; Huettig, Rommers, & Meyer, 2011). Recently, researchers have therefore proposed more sophisticated statistical techniques such as multi-level logistic regression (Barr, 2008), and Growth Curve Analyses (Mirman et al., 2008, 2014). In the current research, following Mirman et al. (2008), the Growth Curve Models were used to analyse participants' fixation distributions.

The proportions of fixations on each picture along the time sequence of the sentences were calculated for the three conditions. A sample was taken every 50 milliseconds (ms), and the nominal phrase time window was chosen from 400 ms before the onset of the nominal phrases (i.e. the numeral), to 400 ms after the end of the nominal phrases (i.e. the offset of the nouns). As mentioned before, in each audio sentence, the normed duration of the numeral is 300ms, the classifier is 250ms, the adjective is 300ms, and the noun is 350ms. In the following part of this section, the duration for each item in the nominal phrases are presented by rectangles with different colours: the grey rectangle represents the duration of the numeral, the blue rectangle represents the duration of the classifier, the red rectangle represents the duration of the adjective, and the green rectangle represents the duration of the noun. Data from each trial were averaged across items and participants to yield continuous time-course estimates of the fixations of the four pictures. Since the picture type was manipulated as a within-participant factor, the overall time course of fixation proportions was modelled with a third-order (cubic) orthogonal polynomial and fixed effects of the picture type on all time terms. The model also included participant random effects on all time terms and participant-by-type random effects on all time terms except the cubic (estimating random effects is “expensive” in terms of the number of observation required, so this cubic term is excluded because it tends to capture less-relevant effects in the tails) (Mirman et al, 2008, 2014).

An analysis time window of interest is defined that extended from 200 ms following the mean onset of a word to 200ms following the mean offset of this word (i.e. the mean



onset of the next word). This 200 ms buffer following the onset of a word is based on the mean time required to plan and launch an eye movement, and the typical lag observed between eye movements and fine-grained phonetic detail in the speech stream (Allopenna, Magnuson, & Tanenhaus, 1998; Kukona, Fang, Aicher, Chen, & Magnuson, 2011). The analysis window of different items is represented by rectangles with different types of lines: the rectangle with the dotted lines represents the analysis window of the classifier, the rectangle with the two-dashed lines represents the analysis window of the adjective, and the rectangle with the dashed lines represents the analysis window of the noun.

To take a close look at how natives Mandarin speakers and high proficiency late Dutch-Mandarin learners on-line take advantage of mass/count structures to build anticipations, the fixation proportions in each picture in three conditions are reported separately. In addition, the analysis window of the items occurred before the nouns are focused firstly, since the changes of the proportions of fixation in each picture before the onset of the noun reveal participants' anticipation building before they actually reach the noun, which is a decisive informative cue for the final target picture. Also, the on-line processing of these items should not be affected by the count/mass status of the nouns since participants should have finished processing the classifiers and the adjectives before they reach the nouns. Thus the fixation distributions before the analysis window of the noun were analysed by averaging count nouns and mass nouns.

On reaching the nouns, the mass/count status of nouns are expected to have an influence on participants' fixation distributions. Also, participants' fixation proportions in the four types of pictures should correlate with their final choices. Thus, participants' fixation distributions in the analysis window of the nouns are discussed later by separating count nouns from mass nouns, and with participants' behaviour data, which illustrating their final choice of the target pictures.

The Growth Curve Model analysis of the fixation proportions in each interest time window (the classifier and the adjective) are reported (the data in the analysis window of the noun will be reported later). The statistical procedure was conducted using the *lmer* program (lme4 package; Bates, Maechler, & Dai, 2008) in the R system (R Development Core Team, 2008). The key indexes include the estimate (*Estimate*),

standard errors ( $SE$ ), and the corresponding  $p$  values ( $p$ ). All the figures in the current study were generated with ggplot2 (Wickham, 2009).

### *5.1.1. Condition 1 with the structure [Num+CL+N]*

In Condition 1 where the nominal phrases have the structure [Num+CL+N], there is only one interest window before the noun: the classifier. The proportions of fixations in the four pictures for native Mandarin speakers and late Dutch-Mandarin learners are plotted in Figure 1.

In Figure 1, the top plot presents the fixation distribution patterns of native Mandarin speakers, while the bottom one presents the fixation distribution patterns of late Dutch-Mandarin learners. Rectangles with different colours represent the duration (from the onset to the offset) of each item, and rectangles with different types of lines represent the analysis windows of different items. Points and lines with different colours represent the proportion of fixations in different pictures: red  $\rightarrow$  the mass-high-Adj picture (i.e. the mass-noun denoting picture in which the size of the group/unit is consistent with the adjective in the nominal phrase), green  $\rightarrow$  the count-count picture (i.e. the count-noun denoting picture in which the objects are presented in individual natural units), blue  $\rightarrow$  the count-mass picture (i.e. the count-noun denoting picture in which the objects are organized in groups or divided in parts), purple  $\rightarrow$  the mass-low-Adj picture (i.e. the mass-noun denoting picture in which the size of the group/unit is inconsistent with the adjective in the nominal phrase).

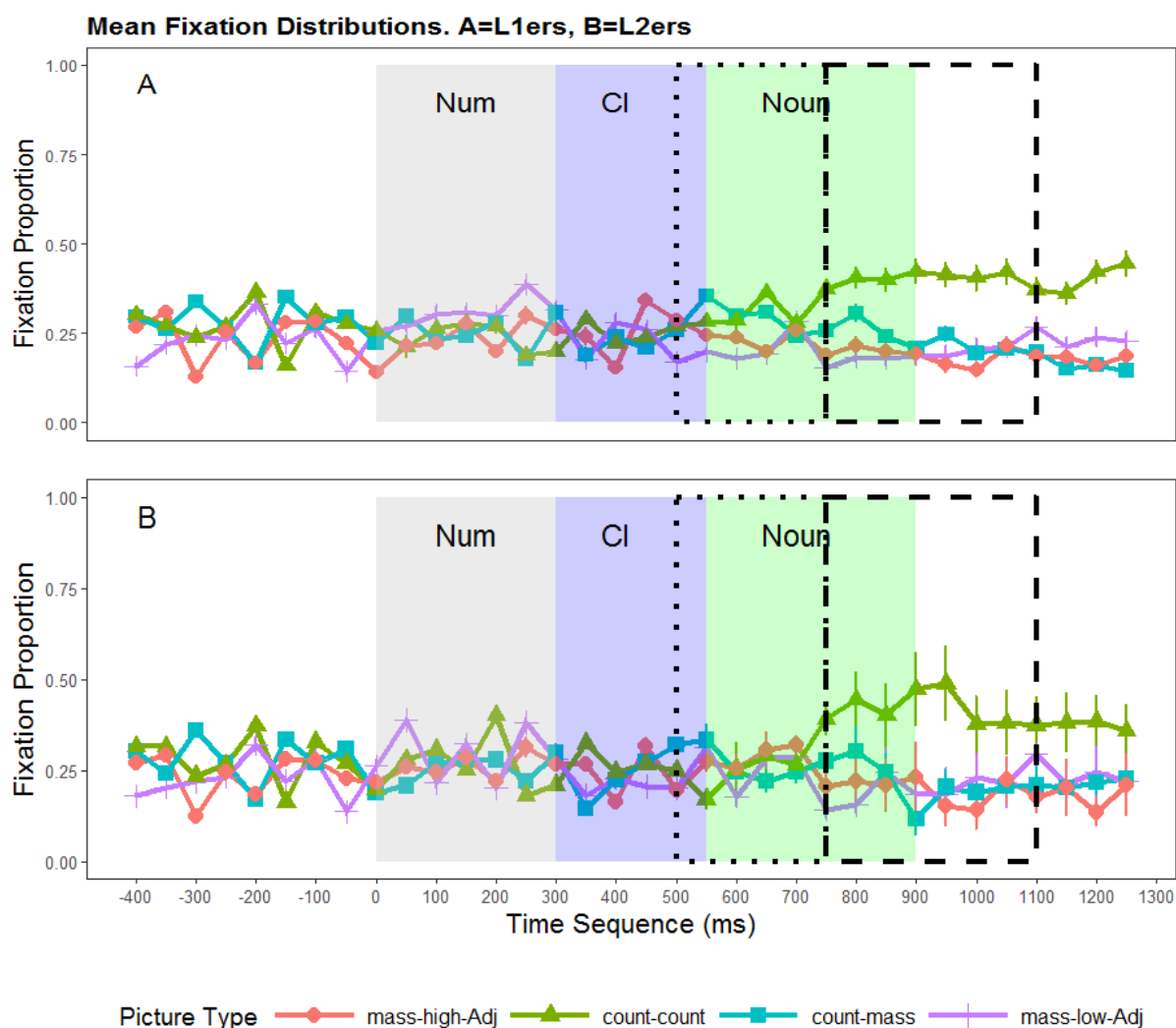


Figure 1. Fixation distributions in Condition 1

From the plots in Figure 1 we can see that, for both native Mandarin speakers (L1ers) and late Dutch-Mandarin learners (L2ers), the proportions of fixations in four pictures stayed around the random level (0.25) before they reach the noun. There is no obvious separation of the fixation distributions in the four pictures until the onset of the noun. This is consistent with the prediction that in the neutral structure [Num+CL+N], there is no informative cue which can be used to build anticipations before the onset of the noun. Both L1ers and L2ers can only decide which one is the target picture on hearing the noun. The proportions of fixations in the interest window of the noun (represented by the rectangle with the dashed lines) revealed a clear division between the count-count picture and the other three pictures. It should be noted that the fixation proportions displayed in Figure 1 are the results after averaging count nouns and mass nouns. It is surprising to

see that in this case the count-count picture still attracted more fixations than the other three pictures. This unexpected finding will be discussed later with participants' behaviour data.

The Growth Curve Model results showed that, in the analysis window of the classifier (represented by the rectangle with the dotted lines), the effect of the picture type on the proportion of fixation is not significant for either L1ers or L2ers, indicating that the onset of the classifier in Condition 1 did not drive participants' fixations significantly more to any one of the four pictures. The detailed results of the Growth Curve Model are illustrated in Table 9.

*Table 9. Growth Curve Model results in Condition1 in the analysis window of the classifier*

	L1ers			L2ers		
	<i>Estimate</i>	<i>SE</i>	<i>p</i>	<i>Estimate</i>	<i>SE</i>	<i>p</i>
<b>Intercept</b>	0.222	0.014	<0.001	0.207	0.021	<0.001
<b>count-count</b>	0.031	0.019	0.264	0.015	0.025	0.811
<b>count-mass</b>	0.024	0.019	0.378	0.011	0.024	0.785
<b>mass-low-Adj</b>	0.013	0.02	0.52	0.007	0.024	0.702

In the Growth Curve Model, the proportion of fixations in the mass-high-Adj picture was chosen as the base line, the proportion of fixations in the other three pictures were analysed by comparing to the mass-high-Adj picture. The data in Table 7 indicates that in the analysis window of the classifier in Condition 1, the fixation proportion differences between the count-count picture and the mass-high-Adj picture, the count-mass picture and the mass-high-Adj picture, the mass-low-Adj picture and the mass-high-Adj picture are not significant.

### **5.1.2. Condition 2 with the structure [Num+CL+Adj+N]**

In Condition 2 where the nominal phrases have the structure [Num+CL+Adj+N], there are two interest windows before the noun: the classifier and the adjective. The proportions of fixations in the four pictures for native Mandarin speakers and late Dutch-Mandarin learners are plotted in Figure 2.

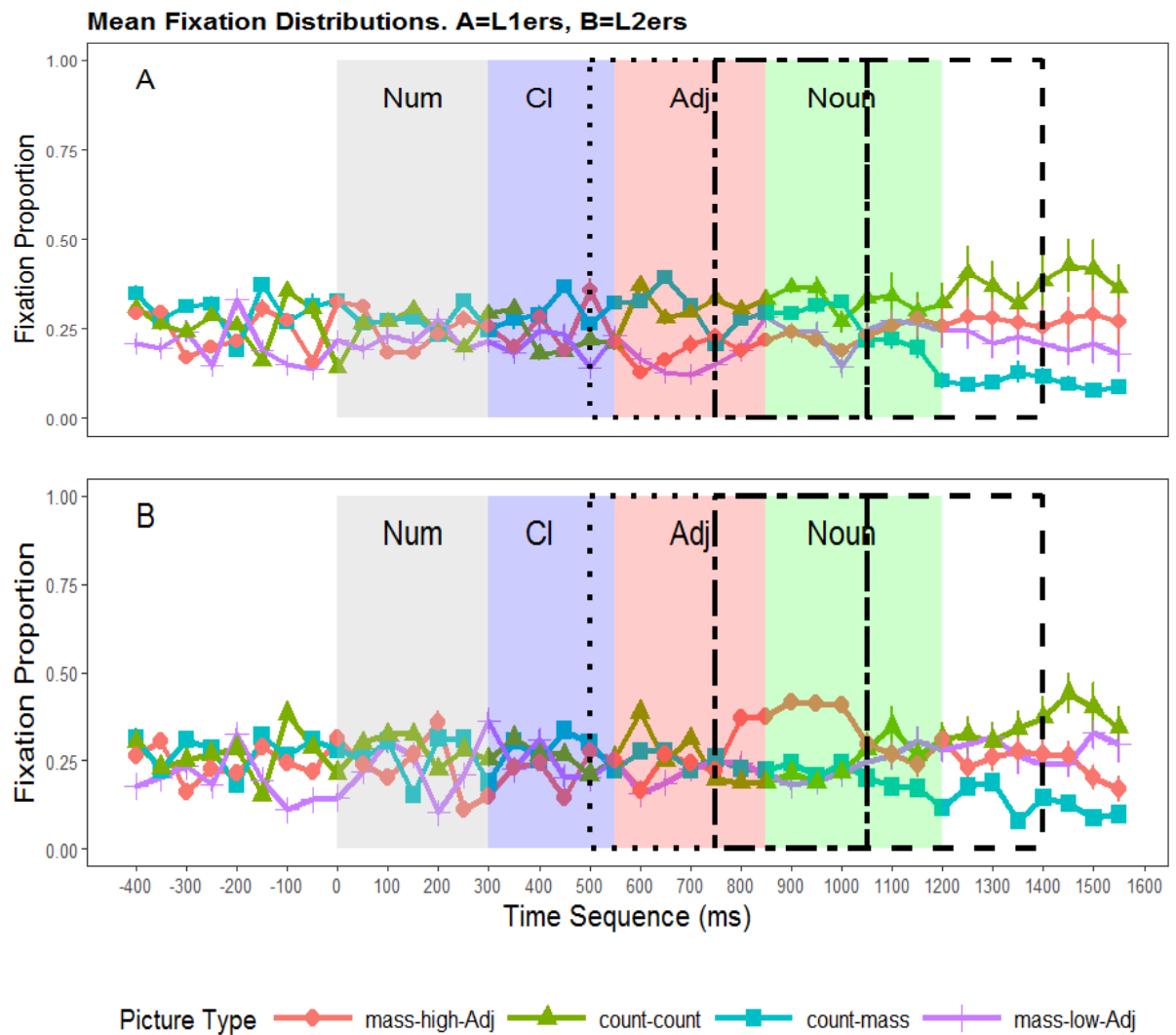


Figure 2. Fixation distributions in Condition 2

From the plots in Figure 2 we can see that, similar to the fixation distribution patterns in Condition 1 that both L1ers and L2ers randomly stared at the four pictures on hearing the classifiers, in Condition 2, the onset of the classifiers did not evoke any preference for any pictures. However, different from Condition 1, in Condition 2, L1ers and L2ers exhibited different patterns on hearing the adjectives. The onset of the adjectives following the classifiers did not affect L1ers' fixation distributions: the proportion of fixations in four pictures stay around the random level in the analysis window of the adjective. However, on hearing the adjectives, L2ers started to land more fixations to the mass-high-Adj picture than the other three pictures, indicating that the onset of the

adjectives attracted L2ers' attention to the mass-high-Adj picture in which the sizes of the group/divided units are consistent with the adjectives.

The Growth Curve Model results showed that, in the analysis window of the classifier, the effect of the picture type on the proportion of fixation is not significant for either L1ers or L2ers. However, in the analysis window of the adjective (represented by the rectangle with the two-dashed lines), there is a reliable effect of the picture type on the fixation proportions for L2ers but not L1ers. For L2ers, in the analysis window of the adjective, there are significantly more fixations in the mass-high-Adj picture than in the count-count picture, the count-mass picture, and the mass-low-Adj picture. The detailed results of the Growth Curve Model are illustrated in Table 10.

Table 10. Growth Curve Model results in Condition 2 in the analysis window of the classifier and the adjective

	Analysis window of the classifier					
	L1ers			L2ers		
	<i>Estimate</i>	<i>SE</i>	<i>p</i>	<i>Estimate</i>	<i>SE</i>	<i>p</i>
<b>Intercept</b>	0.228	0.016	<0.001	0.317	0.021	<0.001
<b>count-count</b>	0.025	0.024	0.305	0.028	0.027	0.471
<b>count-mass</b>	0.011	0.024	0.639	0.011	0.031	0.788
<b>mass-low-Adj</b>	0.003	0.025	0.934	0.005	0.029	0.945
	Analysis window of the adjective					
	L1ers			L2ers		
	<i>Estimate</i>	<i>SE</i>	<i>p</i>	<i>Estimate</i>	<i>SE</i>	<i>p</i>
<b>Intercept</b>	0.236	0.014	<0.001	0.207	0.021	<0.001
<b>count-count</b>	0.031	0.022	0.151	-0.223	0.023	<b>&lt;0.001</b>
<b>count-mass</b>	0.031	0.022	0.154	-0.207	0.024	<b>&lt;0.001</b>
<b>mass-low-Adj</b>	0.003	0.022	0.896	-0.241	0.023	<b>&lt;0.001</b>

L1ers' fixation distribution pattern is consistent with my prediction that in the nominal phrase with the structure [Num+CL+Adj+N], there is no mass/count syntactic cue which

can offer useful information for building anticipations. Just like in Condition 1, participants need to wait till the onset of the noun to decide which picture is the target. However, L2ers exhibited different patterns as predicted. Their preference for the mass-high-Adj picture in which the size of the group/divided unit is consistent with the adjective on hearing the adjectives indicates that L2ers were sensitivity to the semantic meanings of adjectives, and were easily attracted by the Adj-consistent pictures.

### 5.1.3. Condition 3 with the structure [Num+Adj+CL+N]

In Condition 3 where the nominal phrases have the structure [Num+Adj+CL+N], there are two interest windows before the noun: the adjective and the classifier. The proportions of fixations in the four pictures for native Mandarin speakers and late Dutch-Mandarin learners are plotted in Figure 3.

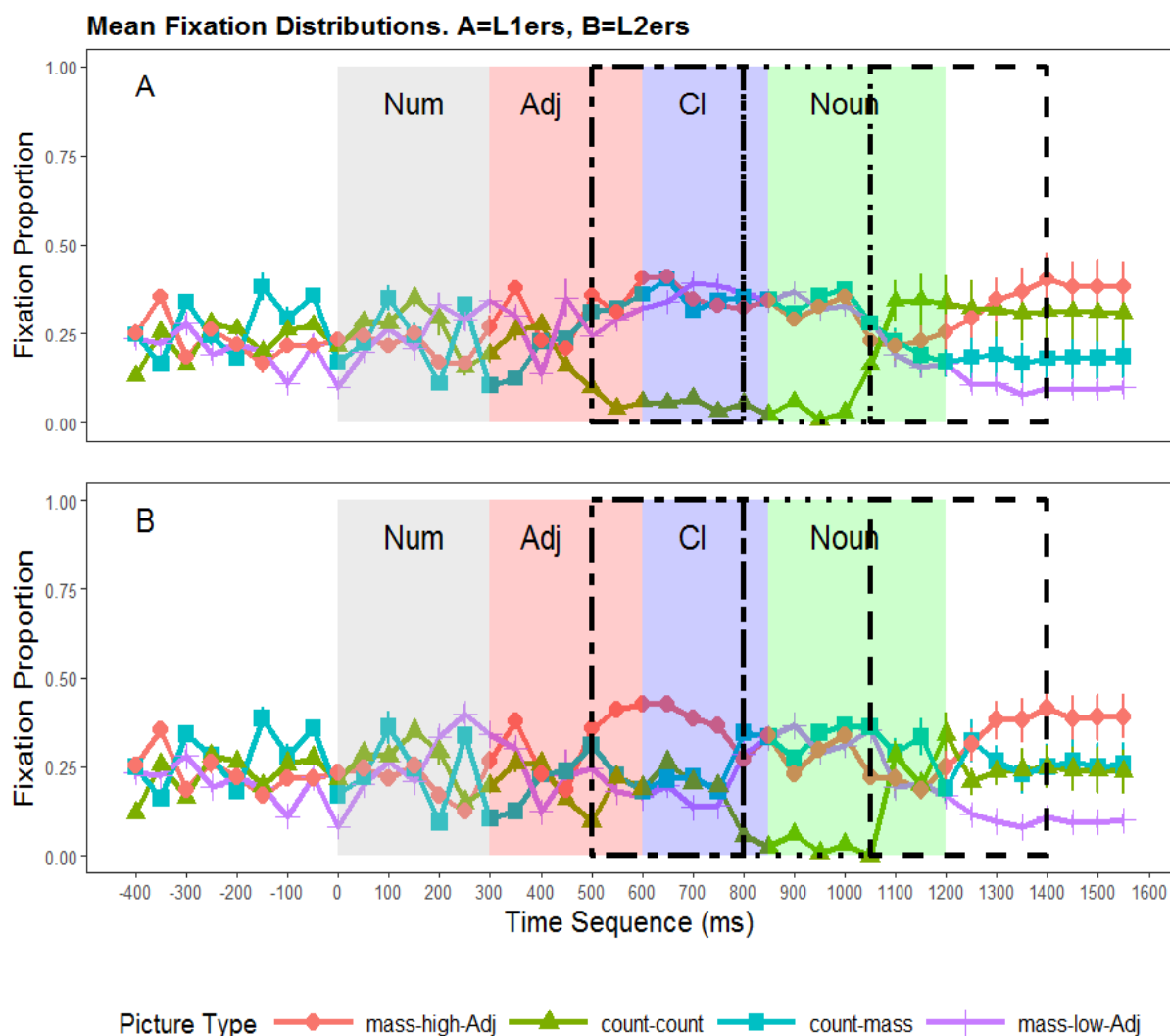


Figure 3. Fixation distributions in Condition 3

From the plots in Figure 3 we can see that, in Condition 3, L1ers and L2ers behaved differently on hearing the adjectives directly following the numeral. For L1ers, the onset of the adjectives drove their attention to the mass/plural-expressing pictures (i.e. the two mass-noun denoting pictures and the count-mass picture). While for L2ers, similar to their patterns in Condition 2, on hearing the adjectives, they directed their fixations to the mass-high-Adj picture in which the size of the group/divided unit is consistent with the adjectives. In addition, the classifiers following the adjectives triggered different patterns on L1ers and L2ers. For L1ers, the onset of the classifiers following the adjectives did not affect their attention, their fixation distributions stayed in the same pattern as in the analysis window of the classifier. For L2ers, however, the onset of the classifiers following the adjectives triggered their attention to the mass/plural-expressing picture (i.e. the two mass-noun denoting pictures and the count-mass picture). Even though the onset of the classifiers following the adjectives affected L1ers and L2ers differently, they both exhibited the preference for the mass/plural-expressing pictures in the analysis window of the classifier, indicating that on hearing the Adj-CL sequence, both L1ers and L2ers built anticipations for a mass/plural interpretation. L1ers' fixation distribution pattern is consistent with my prediction that the Adj-CL word order which is a mass-preferred syntactic cue would force a mass/plural anticipation. On hearing the adjectives directly following the numeral, L1ers already made a prediction about the upcoming item to be a massifier, and the nominal phrase to have a mass/plural interpretation. However, L2ers' fixation distribution pattern is different from my prediction. The adjectives directly following the numeral did not trigger their expectations for a massifier or a mass/plural interpretation. Instead, it drove their attention to the picture in which the objects were organized in an Adj-consistent size. This Adj-sensitive pattern of L2ers has also been found in Condition 2. On hearing the classifiers following the adjectives, L2ers started to shift their attention to the mass/plural-expressing pictures, which is a sign of being sensitive to the mass/count structures based on my prediction.

The Growth Curve Model results revealed a reliable effect of the picture type on both L1ers' and L2ers' fixation distributions in the analysis window of both the adjective and the classifier. In the analysis window of the adjective, L1ers landed significantly more fixations in the mass-high-Adj picture than in the count-count picture. While the



difference between the mass-high-Adj picture and the count-mass picture, and the difference between the mass-high-Adj picture and the mass-low-picture are not significant. On the other hand, L2ers landed significantly more fixations in the mass-high-Adj picture than the other three pictures. In the analysis window of the classifier, both L1ers and L2ers landed more fixations in the mass-high-Adj picture than in the count-count picture. The difference of fixation proportions between the mass-high-Adj picture and the count-mass picture, and between the mass-high-Adj picture and the mass-low-Adj picture are not significant. The detailed results of the Growth Curve Model are illustrated in Table 11.

Table 11. Growth Curve Model results in Condition 3 in analysis window of the adjective and the classifier

	Analysis window of the adjective					
	L1ers			L2ers		
	<i>Estimate</i>	<i>SE</i>	<i>p</i>	<i>Estimate</i>	<i>SE</i>	<i>p</i>
<b>Intercept</b>	0.228	0.016	<0.001	0.317	0.021	<0.001
<b>count-count</b>	-0.358	0.021	<b>&lt;0.001</b>	-0.282	0.022	<b>&lt;0.001</b>
<b>count-mass</b>	0.007	0.024	0.839	-0.265	0.023	<b>&lt;0.001</b>
<b>mass-low-Adj</b>	0.004	0.025	0.901	-0.218	0.023	<b>&lt;0.001</b>
	Analysis window of the classifier					
	L1ers			L2ers		
	<i>Estimate</i>	<i>SE</i>	<i>p</i>	<i>Estimate</i>	<i>SE</i>	<i>p</i>
<b>Intercept</b>	0.236	0.014	<0.001	0.207	0.021	<0.001
<b>count-count</b>	-0.313	0.022	<b>&lt;0.001</b>	-0.301	0.028	<b>&lt;0.001</b>
<b>count-mass</b>	0.005	0.022	0.854	-0.002	0.028	0.957
<b>mass-low-Adj</b>	0.003	0.022	0.896	-0.003	0.028	0.941

The results of L1ers' and L2ers' fixation distribution patterns showed that different nominal phrase structures affected participants' on-line processing and anticipations building. Both L1ers and L2ers were sensitive to the Adj-CL word order which is a mass-preferred syntactic cue based on Cheng & Sybesma (1998, 1999). On hearing the Adj-

CL structure in Condition 3, both L1ers and L2ers directed their attention to the mass/plural-expressing pictures (the two mass-noun denoting pictures and the count-mass picture) instead of the count-expressing picture (the count-count picture). Different from L1ers, L2ers exhibited their sensitivity to the lexical-semantic information of the adjectives: in both Condition 2 and Condition 3, they were easily attracted by the adjectives and the pictures containing objects grouped in an Adj-consistent size.

The fixation distributions before the onset of the nouns revealed participants' predictive processing of the nominal phrases with different structures before they reach the nouns. The results showed that L1ers and L2ers built different anticipations on hearing the classifiers and the adjectives in different structures. In order to further look into how they responded to the mass/count nouns, and how they integrated the anticipations built based on the structures and the actual nouns they heard, participants' fixation proportions in the analysis window of the noun and their behaviour data were analysed together.

## **5.2. Behavioural data and fixation distributions on the nouns**

### **5.2.1. Behavioural data**

In analysing participants' behaviour data, their choice of the target picture based on different nominal phrases structures and count/mass nouns were calculated. Considering all the factors are categorical variables (nominal phrase conditions, mass/count nouns, picture types), the Generalized Linear Model (GLM) was used to explore the effect of the nominal phrase structures and the mass/count nouns on participants' picture choosing. By adding random effects from participants and items in the model, the GLM approach is more powerful than the traditional ANOVA approach with separate participant (F1) and item (F2) analyses. The statistical procedure was conducted using the *glm* program (glm package; R Core team, 2008) in the R system (R Development Core Team, 2008).

When using the GLM, there is no consensus on which is the best way to do the data analysis: confirmatory or exploratory. In order to examine the effects of the factors manipulated in the current study (the nominal phrase structures and the mass/count nouns), and at the same time, explore the potential influences from other possible factors, both confirmatory and exploratory analysis were used. For both L1ers and L2ers, two GLM models were applied to the behaviour data: the confirmatory model and the explored best fit model. In the confirmatory model, the nominal phrase structures (3

levels) and the mass/count status of nouns (2 levels) are the two fixed factors, and participants and items are random factors. In the explored best fit model, I started with a ‘full model’ (e.g. Zuur, Ieno, Walker, Saveliev, & Smith, 2009; Barr, Levy, Scheepers, & Tily, 2013) with all the possible factors – the nominal phrase structures (3 levels), the mass/count nouns (2 levels), and the classifiers (4 levels) – as the three fixed factors, participants and items as the two random factors. Based on this ‘full model’, using the step-wise regression, I reduced the model by systematically removing non-significant terms, and eventually got the best fit model (Yan, Zhou, Shu, Yusupu, Miao, Krugel, & Kliegl, 2014). In the following, the results of both the confirmatory model and the explored best fit model are presented for both native Mandarin speakers (L1ers) and late Dutch-Mandarin learners (L2ers), reporting the estimate (*Estimate*), standard errors (*SE*), Z values (*Z*) and the corresponding *p* values (*p*). All the figures in the current study were generated with ggplot2 (Wickham, 2009).

The results of the confirmatory model showed that neither the nominal phrase structures nor the mass/count nouns affected L1ers’ final choice of the target pictures: *Estimate* = 1.957e+01, *SE* = 9.817e+02, *Z* = 0.02, *p* = 0.984; *Estimate* = -2.008e+01, *SE* = 9.817e+02, *Z* = -0.02, *p* = 0.982. Similarly, neither the nominal phrase structures nor the mass/count nouns significantly affected L2ers’ final choice of the target pictures: *Estimate* = 1.506e+01, *SE* = 3.965e+03, *Z* = 0.01, *p* = 0.998; *Estimate* = 1.828e+01, *SE* = 3.965e+03, *Z* = 0.01, *p* = 0.989.

By using the step-wise regression, the best fit model for L1ers and L2ers are same, which is illustrated in (35).

(35) Picture ~ nominal structures+ mass/count noun+ classifier+ nominal  
structures:classifier

The results of the best fit model in (31) exhibited different results to the confirmatory model. In the best fit model, both the nominal phrase structures and the mass/count nouns significantly affected participants’ final choice. And apart from these two manipulated factors, different classifiers also affected their final choice. In addition, there exist significant interactions between the nominal phrase structures and different classifiers. This is consistent with my predictions summarized in Table 7. The results of the best fit model for both L1ers and L2ers are summarized in Table 12. The detailed results are presented in Appendix B.1.

Table 12. The best fit model results of behaviour data

<i>Diff</i>	<b>L1ers</b>	<b>L2ers</b>
Nominal C2-C1	1.1987*	2.0711***
Nominal C3-C2	-2.8444***	-1.3972*
Nominal C3-C2	-4.0431***	-3.5573***
Noun M-C	-21.8908*	-19.1582*
Classifier <i>gen-ba</i>	-6.3e-01	-5.8e-03
Classifier <i>kuai-ba</i>	-6.3e-01	-5.8e-01
Classifier <i>kuai-gen</i>	-2.9e-15	-2.9e-07
Condition 2: Classifier <i>gen</i>	-1.5664*	-1.2469*
Condition 3: Classifier <i>gen</i>	2.1889*	1.5976*
Condition 2: Classifier <i>kuai</i>	-1.5624*	-1.3428*
Condition 3: Classifier <i>kuai</i>	2.1631	1.7998*

\*\*\*  $p < 0.001$ , \*  $p < 0.05$

To further explore the interactions between the nominal structures and different classifiers, the Tukey post-hoc tests were conducted by separating different classifiers. When the classifier was *ba*, L1ers and L2ers have the same best fit model, which is illustrated in (36).

(36) Picture ~ nominal structures + mass/count noun

From the model in (36) we can see that only the nominal structures and count/mass nouns affected L1ers' and L2ers' final choice of the target pictures when the classifier was *ba*, but not their interactions. The results are summarized in Table 13. The detailed results are presented in Appendix B.2.

Table 13. The results of Tukey post-hoc test on behavioural data when the classifier is *ba*

	<b>L1ers</b>	<b>L2ers</b>
Nominal C2-C1	1.198*	1.277*
Nominal C3-C1	-2.844***	-2.895***
Nominal C3-C2	-4.043***	-4.431***
Noun M-C	-22.21***	-21.76***

\*\*\*  $p < 0.001$ , \*  $p < 0.05$

The best fit models for both of the groups for the classifier *gen* and the classifier *kuai* are the same, as showed in (37).

(37) Picture ~ mass/count noun

From the best fit model in (37) we can see that when the classifier was *gen/kuai*, the only factor that affected L1ers' and L2ers' final choice is the mass/count status of nouns. Different nominal phrase structures did not affect their final choice. The results are summarized in Table 14. The detailed results are presented in Appendix B.3.

Table 14. The results of Tukey post-hoc test on behavioural data when the classifier is *gen/kuai*

	<b>L1ers</b>	<b>L2ers</b>
<i>Gen</i> : M-C	-20.3***	-18.03***
<i>Kuai</i> : M-C	-20.62***	-18.84***

\*\*\*  $p < 0.001$

The GLM best fit models showed that apart from the two manipulated factors (nominal phrase structures and the mass/count nouns), different classifiers also affected L1ers' and L2ers' final interpretations. L1ers and L2ers exhibited similar patterns in their behavioural data. The classifier *ba* evoked different patterns from the classifier *gen* and *kuai*, which is an interesting finding and will be discussed later.

Both L1ers' and L2ers' final choices of the target pictures on hearing nominal phrases with different structures are illustrated in Table 15.

Table 15. L1ers' and L2ers' target picture choices

<i>ba</i>	Native Mandarin speakers						Late Dutch-Mandarin learners					
	Condition 1		Condition 2		Condition 3		Condition 1		Condition 2		Condition 3	
	Count	Mass	Count	Mass	Count	Mass	Count	Mass	Count	Mass	Count	Mass
<b>M-hi</b>	0 %	<b>51%</b>	0 %	25%	0 %	<b>95%</b>	0%	<b>55%</b>	0%	40%	0%	<b>80%</b>
<b>C-C</b>	<b>80%</b>	0 %	<b>95%</b>	0%	1%	0 %	<b>84%</b>	0%	<b>80%</b>	0%	5%	0%
<b>C-M</b>	20%	0 %	5%	0 %	<b>99%</b>	0 %	16%	0%	20%	0%	<b>95%</b>	0%
<b>M-lo</b>	0%	49%	0 %	<b>75%</b>	0 %	5%	0%	45%	0%	<b>60%</b>	0%	20%
<i>Gen</i> <i>kuai</i>	Condition 1		Condition 2		Condition 3		Condition 1		Condition 2		Condition 3	
	Count	Mass	Count	Mass	Count	Mass	Count	Mass	Count	Mass	Count	Mass
<b>M-hi</b>	0 %	<b>55%</b>	0 %	<b>75%</b>	0 %	<b>80%</b>	0%	50%	0%	<b>80%</b>	0%	<b>60%</b>
<b>C-C</b>	<b>90%</b>	0 %	<b>90%</b>	0 %	<b>95%</b>	0 %	<b>95%</b>	0%	<b>95%</b>	0%	<b>95%</b>	0%
<b>C-M</b>	10%	0 %	10%	0 %	5%	0 %	5%	0%	5%	0%	5%	0%
<b>M-lo</b>	0%	45%	0 %	25%	0 %	20%	0%	50%	0%	20%	0%	40%

(*M-hi* stands for the mass-high-Adj picture, *M-lo* stands for the mass-low-Adj picture, *C-C* stands for the count-count picture, *C-M* stands for the count-mass picture)

From Table 15 we can see a clear separate pattern between the classifier *ba* and the classifiers *gen/kuai* on both native Mandarin speakers and late Dutch-Mandarin learners. When the classifier was *ba*, in Condition 1, most of the natives (80%) chose the count-count picture as the target picture when the noun was a count noun, while half of them chose either the mass-high-Adj picture or the mass-low-Adj picture when the noun was a mass noun. In Condition 2, most of L1ers (95%) chose the count-count picture as the target picture when the noun was a count noun, while most of them (75%) chose the mass-low-Adj picture when the noun was a mass noun. In Condition 3, almost all the natives (99%) chose the count-mass picture when the noun was a count noun, while most of them (95%) chose the mass-high-Adj picture when the noun was a mass noun. As for

L2 learners, in Condition 1, most of them (84%) chose the count-count picture when the noun was a count noun, while 55% of them chose the mass-high-Adj picture when the noun was a mass noun. In Condition 2, 80% of L2ers chose the count-count picture when the noun was a count noun, while 60% of them chose the mass-low-Adj picture when the noun was a mass noun. In Condition 3, 95% of them chose the count-mass picture when the noun was a count noun, while 80% of them chose the mass-high-Adj picture when the noun was a mass noun.

When the classifier was *gen/kuai*, in Condition 1, most of the natives (90%) chose the count-count picture as the target picture when the noun was a count noun, while half of them chose either the mass-high-Adj picture or the mass-low-Adj picture when the noun was a mass noun. In Condition 2, most of the natives (90%) chose the count-count picture when the noun was a count noun, while most of them (75%) chose the mass-high-Adj picture when the noun was a mass noun. In Condition 3, most of the natives (95%) chose the count-count picture when the noun was a count noun, while most of them (80%) chose the mass-high-Adj picture when the noun was a mass noun. As for L2 learners, in Condition 1, 95% of them chose the count-count picture when the noun was a count noun, while either the mass-high-Adj picture or the mass-low-Adj picture when the noun was a mass noun. In Condition 2, 95% of them chose the count-count picture when the noun was a count noun, while 80% of them chose the mass-high-Adj picture when the noun was a mass noun. In Condition 3, 95% of them chose the count-count picture when the noun was a count noun, while 60% of them chose the mass-high-Adj picture when the noun was a mass noun.

From the data in Table 13 and the results of the GLM, it is obvious that L2 learners exhibited the native-like behaviour patterns, indicating that they interpreted the nominal phrases with different structures the same way as natives Mandarin speakers did. By comparing the classifier *ba* to the classifiers *gen/kuai* we can see that both L1ers and L2ers behaved differently to these two groups of classifiers. When the noun in the nominal phrase was a count noun, even though both the count-count picture and the count-mass picture contain objects denoted by the same count noun, participants chose the count-count picture as the target picture in both Condition 1 and Condition 2 when the classifier was *ba*, while choosing the count-mass picture in Condition 3. On the other hand, when the classifier was *gen/kuai*, participants chose the count-count picture as the target picture in all three conditions.

First of all, these results can be used to explain the surprising finding observed in Figure 1: in Condition 1, by averaging count nouns and mass nouns, participants still prefer the count-count picture than the other three pictures. Recall that, no matter what the classifier was (either *ba* or *gen* or *kuai*), in Condition 1, most of the participants chose the count-count picture as the target than the count-mass picture when the noun was a count noun (85% vs. 15%), while half of them chose the mass-high-Adj picture and half of them chose the mass-low-Adj picture when the noun was a mass noun (50% vs. 50%). In this case, when averaging count nouns and mass nouns, the count-count picture still has the highest possibility to be chosen as the target among the four pictures in Condition 1. As a consequence, it is reasonable to observe that the count-count picture gained more fixations than the other three pictures in Figure 1.

Furthermore, these results also indicate that when the classifier was *ba*, participants' interpretation of these nominal phrases was affected by different structures: they chose different pictures as the target pictures on hearing nominal phrases with and without the mass-preferred syntactic cue. On the contrary, when the classifier was *gen/kuai*, different nominal phrase structures did not affect participants' final interpretations of nominal phrases: they always preferred the count-count picture as the target picture even though heard nominal phrases with the mass syntactic cue (the Adj-CL word order in Condition 3). This is different from my prediction that the mass-preferred structure would always force a mass/plural interpretation. Detailed discussion about this interesting finding is presented in Section 6.

Considering the findings that different classifiers affected participants' interpretations differently, the analysis of participants' fixation distributions in the interest window of the noun were separated by different classifiers and mass/count nouns.

### 5.2.2. Fixation distributions in the analysis window of the noun

#### 5.2.2.1. When the classifier was *ba*

##### 5.2.2.1.1. Condition 1 with the structure [Num+CL+N]

The interest window for the noun in Condition 1 is from 750 ms to 1100 ms, as the rectangle with the dashed lines represented. Since count nouns and mass nouns in the audio materials would evoke different fixation distributions to the four pictures, the fixation proportions were analysed by separating count nouns from mass nouns. Based on the Growth Curve Model, the fitted curves for the fixation proportions in the four



pictures in the count noun window and the mass noun window for both L1ers and L2ers are illustrated in Figure 4.

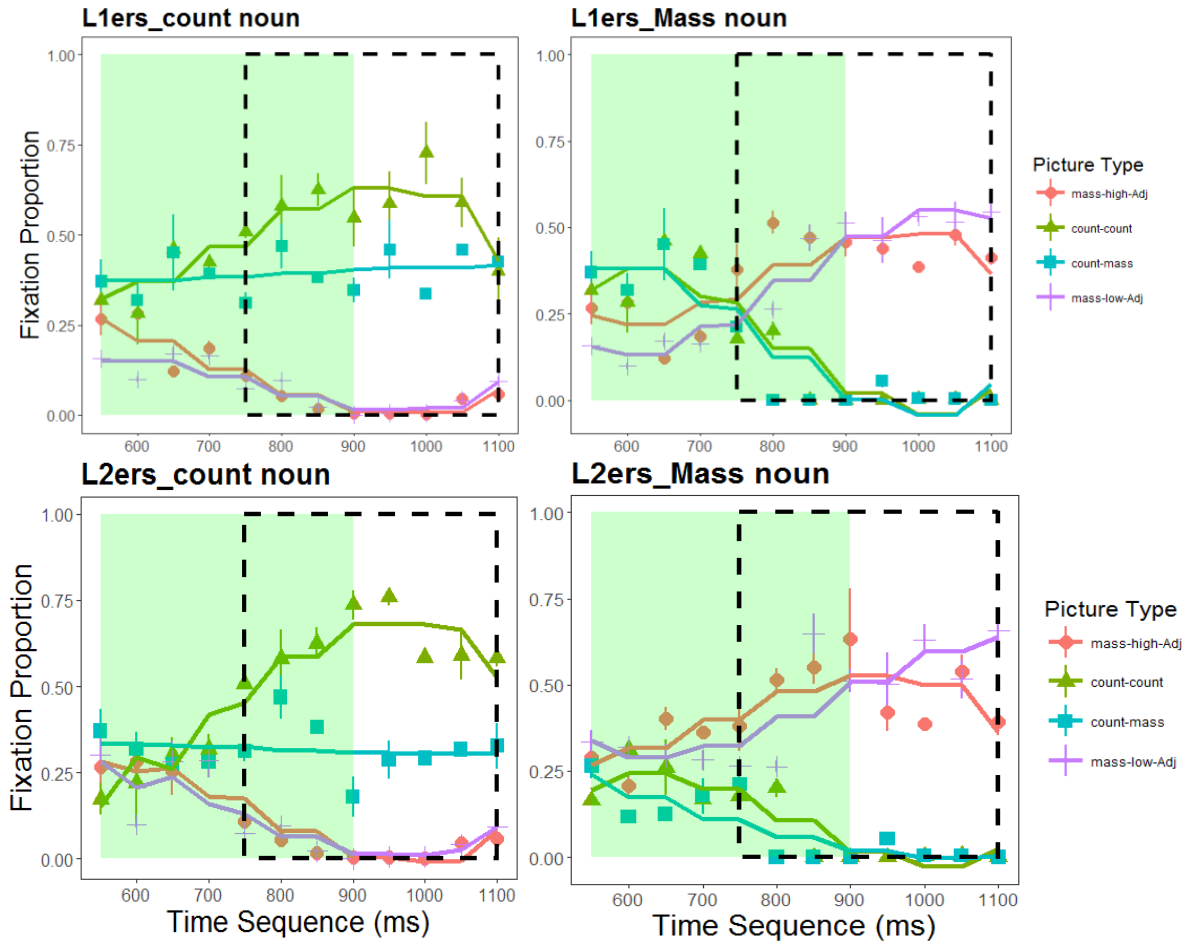


Figure 4. Analysis window of the noun in Condition 1 with ba

In the plots in Figure 4, the points with different colours and shapes represent participants' fixation proportions in each picture, and the lines with different colours represent the fitted curve lines generated based on the Growth Curve Model.

For L1ers, when the noun was a count noun, the picture type had a significant effect on the fixation proportions in the interest window of the noun. There were significantly more fixations in the count-count picture than in the mass-high-Adj picture:  $Estimate = 0.382$ ,  $SE = 0.018$ ,  $p < 0.001$ . Also, more fixations were landed in the count-mass picture than in the mass-high-Adj picture:  $Estimate = 0.284$ ,  $SE = 0.018$ ,  $p < 0.001$ . The difference of fixation proportions between the mass-high-Adj picture and the mass-low-Adj picture was not significant:  $Estimate = -0.023$ ,  $SE = 0.019$ ,  $p = 0.221$ . When the noun was a mass noun, the picture type had a significant effect on the fixation proportions.

There were significantly less fixations in the count-count picture than in the mass-high-Adj picture:  $Estimate = -0.184$ ,  $SE = 0.022$ ,  $p < 0.001$ . Also, significantly less fixations were landed in the count-mass picture than in the mass-high-Adj picture:  $Estimate = -0.183$ ,  $SE = 0.022$ ,  $p < 0.001$ . The difference of fixation proportions between the mass-high-Adj picture and the mass-low-Adj picture was not significant:  $Estimate = -0.008$ ,  $SE = 0.022$ ,  $p = 0.718$ . These observations indicate that the onset of count nouns drove L1ers' attention to the count-noun denoting pictures, while the onset of mass nouns drove their attention to the mass-noun denoting pictures.

As for L2ers, in the interest window of the nouns, they exhibited similar behaviour patterns with L1ers: the onset of count nouns drove L2ers' fixations to the count-noun denoting pictures while the onset of mass nouns drove their attention to the mass-noun denoting pictures. The Growth Curve Model results showed that, for L2ers, when the noun was a count noun, significantly more fixations were landed in the count-count picture than in the mass-high-Adj picture:  $Estimate = 0.513$ ,  $SE = 0.009$ ,  $p < 0.001$ . Also, more fixations were landed in the count-mass picture than in the mass-high-Adj picture:  $Estimate = 0.206$ ,  $SE = 0.014$ ,  $p < 0.001$ . The difference of fixation proportions between the mass-high-Adj picture and the mass-low-Adj picture was not significant:  $Estimate = -0.017$ ,  $SE = 0.026$ ,  $p = 0.283$ . When the noun was a mass noun, significantly less fixations were landed in the count-count picture than in the mass-high-Adj picture:  $Estimate = -0.472$ ,  $SE = 0.006$ ,  $p < 0.001$ . Also, less fixations were landed in the count-mass picture than in the mass-high-Adj picture:  $Estimate = -0.519$ ,  $SE = 0.006$ ,  $p < 0.001$ . The difference of fixation proportions between the mass-high-Adj picture and the mass-low-Adj picture was not significant:  $Estimate = -0.007$ ,  $SE = 0.028$ ,  $p = 0.647$ .

In general, in the interest window of the noun, L2ers exhibited native-like behaviours: the onset of the count nouns led to more fixations on the count-noun denoting pictures, while the onset of the mass nouns led to more fixations on the mass-noun denoting pictures. Also, the fixation distributions among the four pictures are consistent with participants' behaviour results: when the noun was a count noun, most of participants chose the count-count picture as the target picture, while when the noun was a mass noun, half of the participants chose the mass-high-Adj picture and the other half chose the mass-low-Adj picture.

#### 5.2.2.1.2. Condition 2 with the structure [Num+CL+Adj+N]

In Condition 2, the interest window of the noun is from 1050 ms to 1400 ms. The fitted curves of the fixation proportions in this time window for both L1ers and L2ers are illustrated in Figure 5.

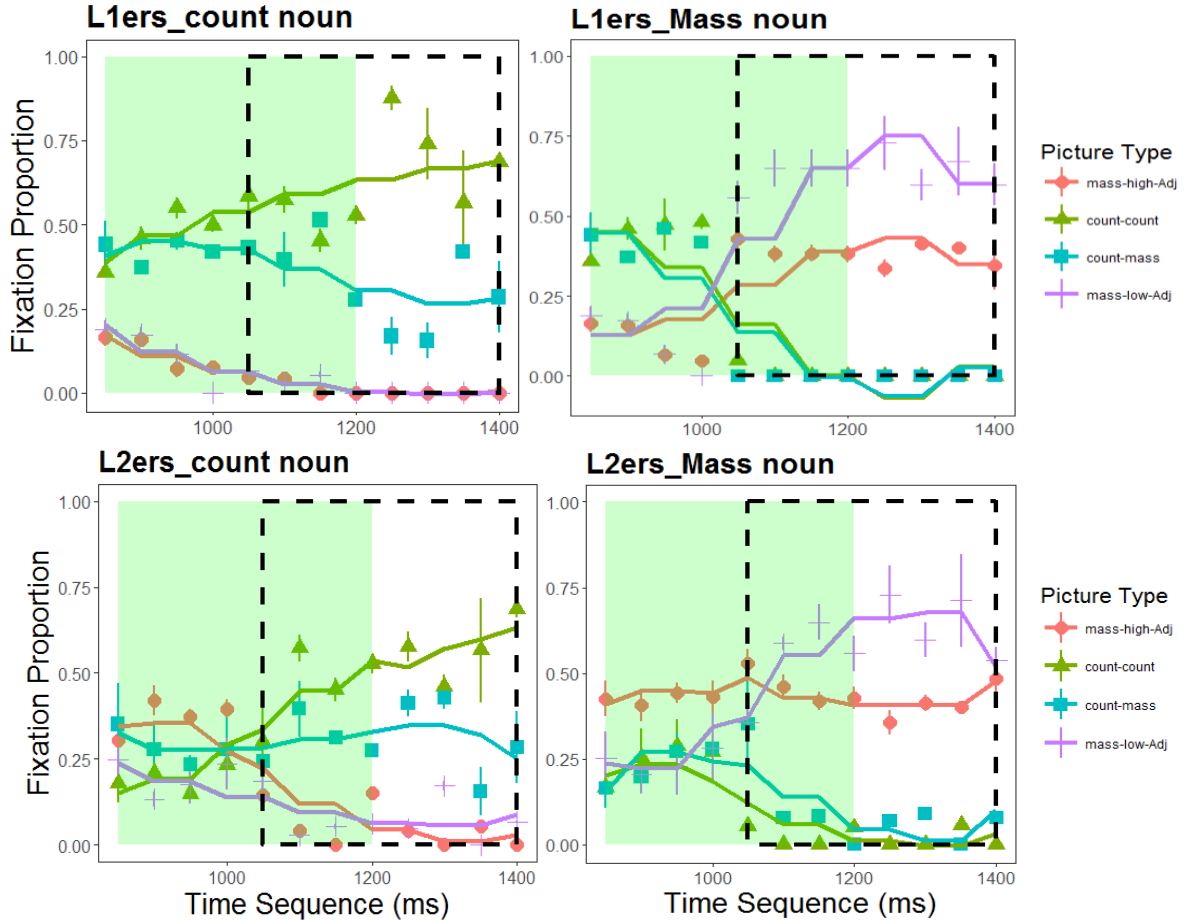


Figure 5. Analysis window of the noun in Condition 2 with *ba*

For L1ers, when the noun was a count noun, the effect of the picture type on the fixation proportion was significant. There were significantly more fixations in the count-count picture than in the mass-high-Adj picture:  $Estimate = 0.516$ ,  $SE = 0.02$ ,  $p < 0.001$ . Also, more fixations were landed in the count-mass picture than in the mass-high-Adj picture:  $Estimate = 0.306$ ,  $SE = 0.02$ ,  $p < 0.001$ . The difference of fixation proportions between the mass-high-Adj picture and the mass-low-Adj picture was not significant:  $Estimate = 0.007$ ,  $SE = 0.02$ ,  $p = 0.734$ . When the noun was a mass noun, the effect of the picture type on the fixation proportion was significant. There were significantly less fixations in the count-count picture than in the mass-high-Adj picture:  $Estimate = -0.141$ ,  $SE = 0.021$ ,

$p < 0.001$ . Also, less fixations were landed in the count-mass picture than in the mass-high-Adj picture:  $Estimate = -0.151$ ,  $SE = 0.021$ ,  $p < 0.001$ . In addition, more fixations were landed in the mass-low-Adj picture than in the mass-high-Adj picture:  $Estimate = 0.169$ ,  $SE = 0.022$ ,  $p < 0.001$ . These observations indicate that count nouns drove participants' attention to the count-noun denoting pictures while mass nouns drove participants' attention to the mass-noun denoting pictures. Recall that the difference between the two mass-noun denoting pictures is in the mass-high-Adj picture, the size of the groups/portions is consistent with the adjective, while in the mass-low-Adj picture, the size of the groups/portions is inconsistent with the adjective, but the size of the individual unit is consistent with the adjective (e.g. the difference between 'one big handful of small pebbles' and 'one small handful of big pebbles'). On hearing the adjectives following the classifiers, L1ers treated the adjectives as modifying the nouns but not the classifiers. Thus they chose the mass-low-Adj picture as the target picture, in which the size of the individual unit of the objects is consistent with the adjectives, but not the size of the groups/portions. This is consistent with their behaviour results that in Condition 2 when the noun was a mass noun, most of L1ers chose the mass-low-Adj picture as the target picture.

For L2ers, when the noun was a count noun, the effect of the picture type on the fixation proportion was significant. There were significantly more fixations in the count-count picture than in the mass-high-Adj picture:  $Estimate = 0.384$ ,  $SE = 0.019$ ,  $p < 0.001$ . Also, more fixations were landed in the count-mass picture than in the mass-high-Adj picture:  $Estimate = 0.317$ ,  $SE = 0.02$ ,  $p < 0.001$ . The difference of fixation proportions between the mass-high-Adj picture and the mass-low-Adj picture was not significant:  $Estimate = 0.004$ ,  $SE = 0.02$ ,  $p = 0.618$ . When the noun was a mass noun, the effect of the picture type on the fixation proportion was significant. There were significantly less fixations in the count-count picture than in the mass-high-Adj picture:  $Estimate = -0.152$ ,  $SE = 0.022$ ,  $p < 0.001$ . Also, less fixations were landed in the count-mass picture than in the mass-high-Adj picture:  $Estimate = -0.141$ ,  $SE = 0.023$ ,  $p < 0.001$ . Moreover, more fixations were landed in the mass-low-Adj picture than in the mass-high-Adj picture:  $Estimate = 0.149$ ,  $SE = 0.018$ ,  $p < 0.001$ . These results show that in the interest window of the nouns L2 exhibited the native-like behaviour pattern.

#### 5.2.2.1.3. Condition 3 with the structure [Num+Adj+CL+N]

In Condition 3 the interest window of the noun is from 1050 ms to 1400 ms. The fitted curves of the fixation proportions in this time window for both L1ers and L2ers are illustrated in Figure 6.

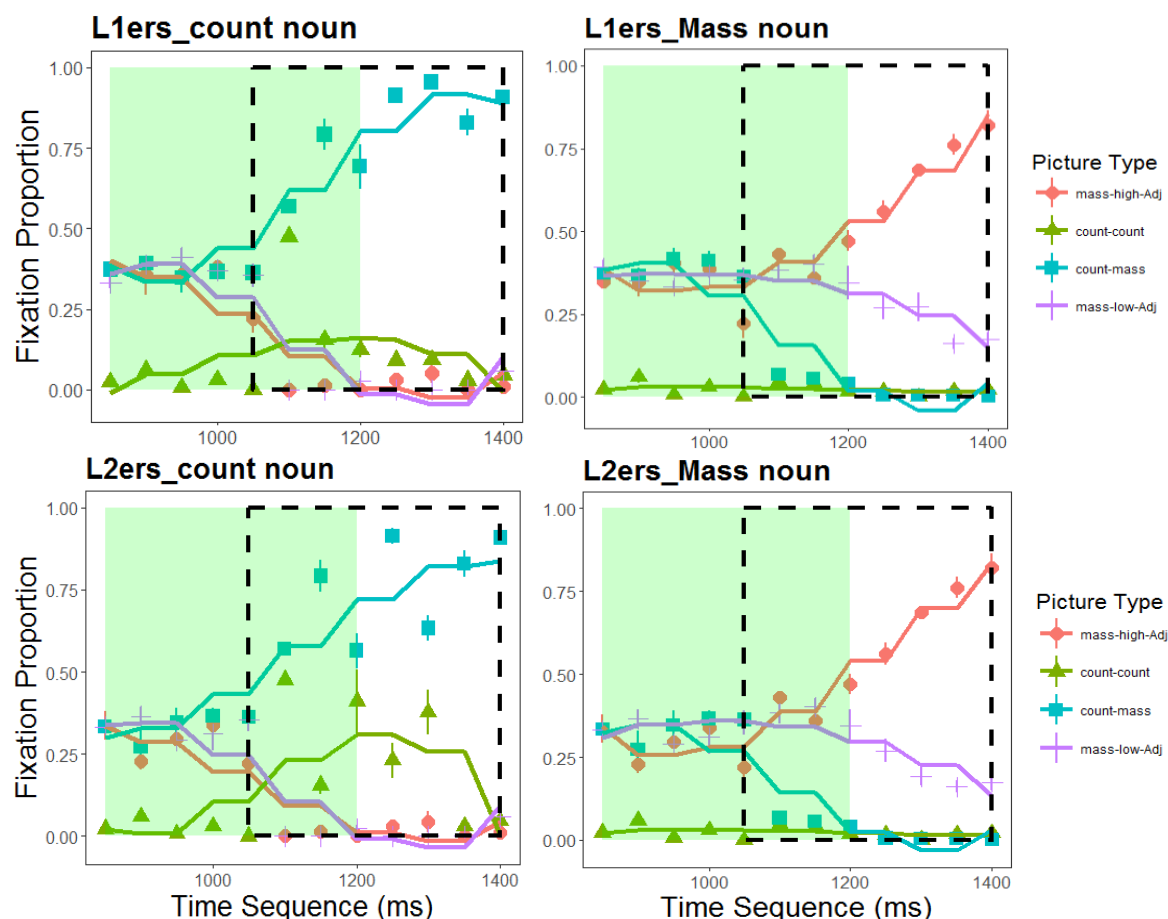


Figure 6. Analysis window of the noun in Condition 3 with ba

For L1ers, when the noun was a count noun, there was a significant effect of the picture type on the fixation proportions. There were significantly more fixations in the count-count picture than in the mass-high-Adj picture:  $Estimate = 0.081$ ,  $SE = 0.018$ ,  $p < 0.001$ . Also, more fixations were landed in the count-mass picture than in the mass-high-Adj picture:  $Estimate = 0.466$ ,  $SE = 0.018$ ,  $p < 0.001$ . The difference of fixation proportions between the mass-high-Adj picture and the mass-low-Adj picture was not significant:  $Estimate = 0.011$ ,  $SE = 0.018$ ,  $p = 0.541$ . When the noun was a mass noun, there was a significant effect of the picture type on the fixation proportions. There were significantly less fixations in the count-count picture than in the mass-high-Adj picture:  $Estimate = -0.479$ ,  $SE = 0.013$ ,  $p < 0.001$ . Also, less fixations were landed in the count-mass picture

than in the mass-high-Adj picture:  $Estimate = -0.321$ ,  $SE = 0.013$ ,  $p < 0.001$ . Besides, less fixations were landed in the mass-low-Adj picture than in the mass-high-Adj picture:  $Estimate = -0.194$ ,  $SE = 0.013$ ,  $p < 0.001$

For L2ers, when the noun was a count noun, there was a significant effect of the picture type on the fixation proportions. There were significantly more fixations in the count-count picture than in the mass-high-Adj picture:  $Estimate = 0.0876$ ,  $SE = 0.022$ ,  $p < 0.001$ . Also, more fixations were landed in the count-mass picture than in the mass-high-Adj picture:  $Estimate = 0.549$ ,  $SE = 0.022$ ,  $p < 0.001$ . The difference of fixation proportions between the mass-high-Adj picture and the mass-low-Adj picture was not significant:  $Estimate = 0.009$ ,  $SE = 0.021$ ,  $p = 0.583$ . When the noun was a mass noun, there was a significant effect of the picture type on the fixation proportions. There were significantly less fixations in the count-count picture than in the mass-high-Adj picture:  $Estimate = -0.542$ ,  $SE = 0.017$ ,  $p < 0.001$ . Also, less fixations were landed in the count-mass picture than in the mass-high-Adj picture:  $Estimate = -0.604$ ,  $SE = 0.017$ ,  $p < 0.001$ . Besides, less fixations were landed in the mass-low-Adj picture than in the mass-high-Adj picture:  $Estimate = -0.187$ ,  $SE = 0.018$ ,  $p < 0.001$ .

By comparing the patterns of L1ers' and L2ers' we can see that L2ers exhibited native-like behaviours in the interest window of the nouns. On hearing the count nouns in Condition 3 which has a mass-biased structure, both of the two groups fixed on the count-mass picture which contains solid objects which were organized in groups. This finding indicates that both L1ers and L2ers interpreted the count nouns in the mass-biased structure to have a mass/plural meaning. Both L1ers and L2ers were sensitive to the mass-biased nominal structure. On hearing the mass nouns, both of the two groups focused on the mass-high-Adj picture in which the size of the group is consistent with the adjective. This finding indicates that both L1ers and L2ers interpreted the adjectives preceding the classifiers to do the job of modifying the classifiers (i.e. the measure/unit) instead of the nouns. The results are consistent with participants' behaviour results.

#### 5.2.2.2. When the classifier was *gen*

##### 5.2.2.2.1. In Condition 1 with the structure [Num+CL+N]

The interest window for the noun in Condition 1 is from 750 ms to 1100 ms, as the rectangle with the dashed lines represented in Figure 7.

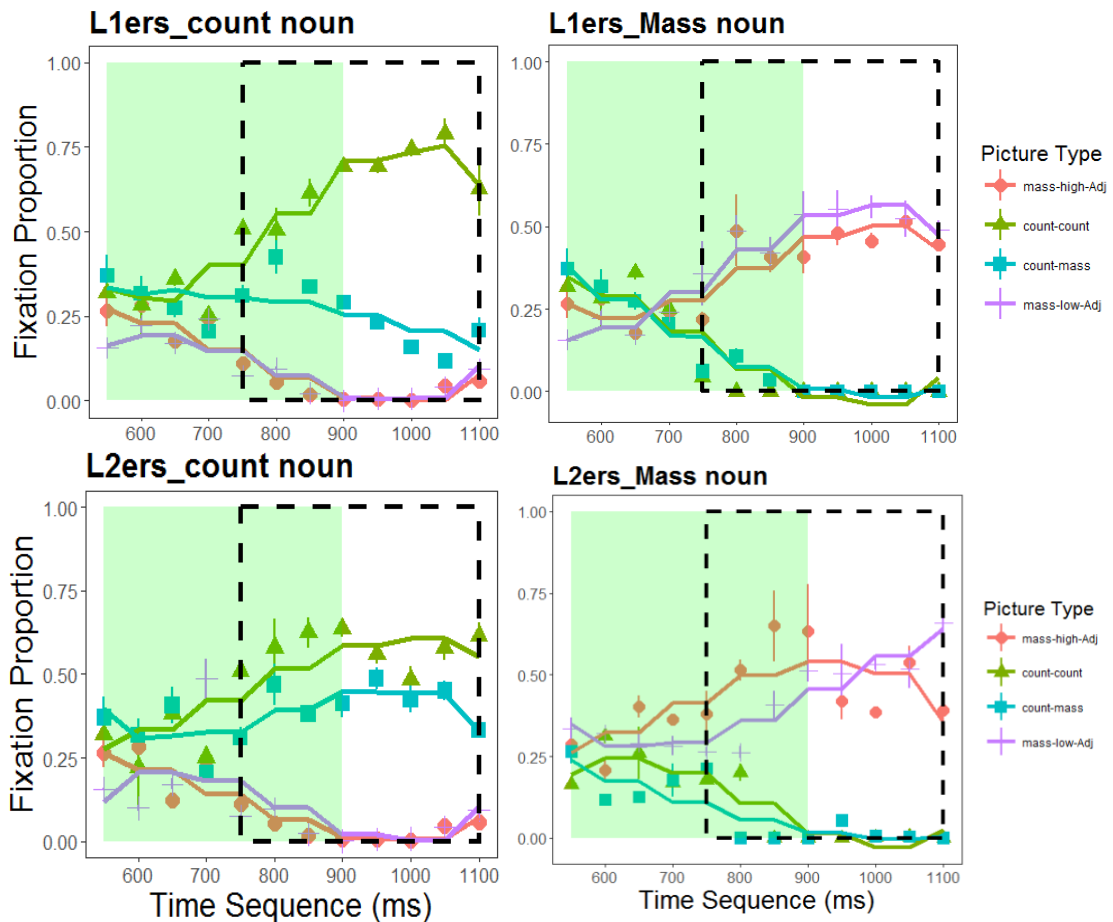


Figure 7. Analysis window of the noun in Condition 1 with gen

For L1ers, when the noun was a count noun, the effect of the picture type on the fixation proportion was significant. There were significantly more fixations in the count-count picture than in the mass-high-Adj picture:  $Estimate = 0.411$ ,  $SE = 0.015$ ,  $p < 0.001$ . Also, more fixations were landed in the count-mass picture than the mass-high-Adj picture:  $Estimate = 0.148$ ,  $SE = 0.015$ ,  $p < 0.001$ . The difference of fixation proportions between the mass-high-Adj picture and the mass-low-Adj picture was not significant:  $Estimate = -0.014$ ,  $SE = 0.015$ ,  $p = 0.362$ . When the noun was a mass noun, the effect of the picture type on the fixation proportion was significant. There were significantly less fixations in the count-count picture than the mass-high-Adj picture:  $Estimate = -0.238$ ,  $SE = 0.018$ ,  $p < 0.001$ . Also, less fixations were landed in count-mass picture than the mass-high-Adj picture:  $Estimate = -0.232$ ,  $SE = 0.018$ ,  $p < 0.001$ . The difference of fixation proportions between the mass-high-Adj picture and the mass-low-Adj picture was not significant:  $Estimate = 0.018$ ,  $SE = 0.018$ ,  $p = 0.332$ . These observations indicate that count nouns drove L1ers' attention to the count noun denoting pictures while mass nouns drove their attention to the mass noun denoting pictures. L2ers exhibited similar patterns to L1ers.

When the noun was a count noun, there were significantly more fixations in the count-count picture than in the mass-high-Adj picture:  $Estimate = 0.372$ ,  $SE = 0.021$ ,  $p < 0.001$ . Also, more fixations were landed in the count-mass picture than the mass-high-Adj picture:  $Estimate = 0.263$ ,  $SE = 0.021$ ,  $p < 0.001$ . The difference of fixation proportions between the mass-high-Adj picture and the mass-low-Adj picture was not significant:  $Estimate = 0.008$ ,  $SE = 0.011$ ,  $p = 0.714$ . When the noun was a mass noun, the effect of the picture type on the fixation proportion was significant. There were significantly less fixations in the count-count picture than the mass-high-Adj picture:  $Estimate = -0.219$ ,  $SE = 0.021$ ,  $p < 0.001$ . Also, less fixations were landed in count-mass picture than the mass-high-Adj picture:  $Estimate = -0.202$ ,  $SE = 0.021$ ,  $p < 0.001$ . The difference of fixation proportions between the mass-high-Adj picture and the mass-low-Adj picture was not significant:  $Estimate = 0.014$ ,  $SE = 0.018$ ,  $p = 0.722$ . These observations indicate that similar to L1ers, count nouns drove L2ers' attention to the count noun denoting pictures while mass nouns drove their attention to the mass noun denoting pictures.

#### 5.2.2.2.2. In Condition 2 with the structure [Num+CL+Adj+N]

The interest window for the noun in Condition 2 is from 1050 ms to 1400 ms, as the rectangle with the dashed lines represented in Figure 8.



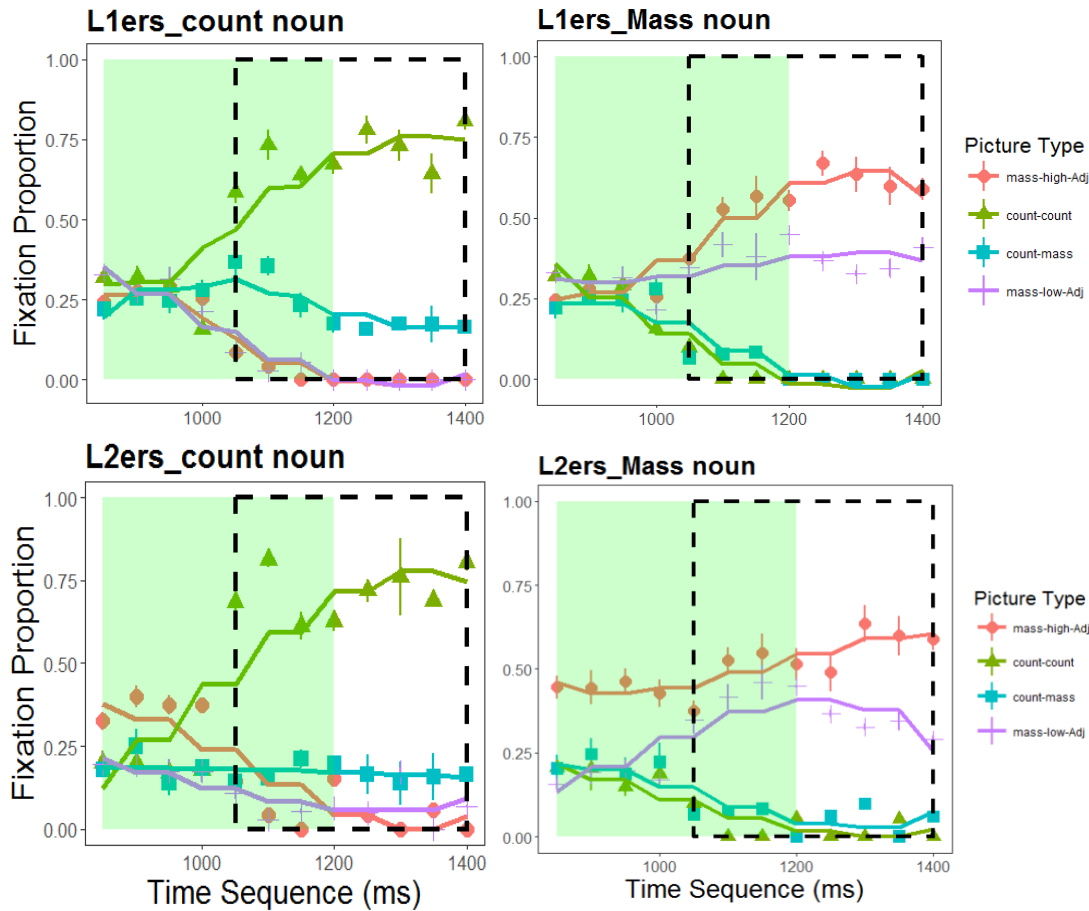


Figure 8. Analysis window of the noun in Condition 2 with gen

For L1ers, when the noun was a count noun, the effect of the picture type on the fixation proportion was significant. There were significantly more fixations in the count-count picture than the mass-high-Adj picture:  $Estimate = 0.449$ ,  $SE = 0.016$ ,  $p < 0.001$ . Also, more fixations were landed in the count-mass picture than the mass-high-Adj picture:  $Estimate = 0.119$ ,  $SE = 0.016$ ,  $p < 0.001$ . The difference of fixation proportions between the mass-high-Adj picture and the mass-low-Adj picture was not significant:  $Estimate = 0.013$ ,  $SE = 0.016$ ,  $p = 0.416$ . When the noun was a mass noun, the effect of the picture type on the fixation proportion was significant. There were significantly less fixations in the count-count picture than the mass-high-Adj picture:  $Estimate = -0.345$ ,  $SE = 0.014$ ,  $p < 0.001$ . Also, less fixations were landed in the count-mass picture than the mass-high-Adj picture:  $Estimate = -0.352$ ,  $SE = 0.014$ ,  $p < 0.001$ . Less fixations were landed in the mass-low-Adj picture than the mass-high-Adj picture:  $Estimate = -0.106$ ,  $SE = 0.017$ ,  $p < 0.01$ . These observations indicate that the onset of count nouns drove L1ers' attention

to the count noun denoting pictures while the onset of mass nouns drove participants' attention to the mass-noun denoting pictures.

L2ers exhibited the same patterns with L1ers. When the noun was a count noun, the effect of the picture type on the fixation proportion was significant. There were significantly more fixations in the count-count picture than the mass-high-Adj picture: *Estimate* = 0.417, *SE* = 0.012,  $p < 0.001$ . Also, more fixations were landed in the count-mass picture than the mass-high-Adj picture: *Estimate* = 0.105, *SE* = 0.012,  $p < 0.001$ . The difference of fixation proportions between the mass-high-Adj picture and the mass-low-Adj picture was not significant: *Estimate* = 0.004, *SE* = 0.011,  $p = 0.654$ . When the noun was a mass noun, the effect of the picture type on the fixation proportion was significant. There were significantly less fixations in the count-count picture than the mass-high-Adj picture: *Estimate* = -0.489, *SE* = 0.021,  $p < 0.001$ . Also, less fixations were landed in the count-mass picture than the mass-high-Adj picture: *Estimate* = -0.413, *SE* = 0.021,  $p < 0.001$ . Less fixations were landed in the mass-low-Adj picture than the mass-high-Adj picture: *Estimate* = -0.201, *SE* = 0.019,  $p < 0.01$ .

#### 5.2.2.2.3. In Condition 3 with the structure [Num+Adj+CL+N]

The interest window for the noun in Condition 3 is from 1050 ms to 1400 ms, as the rectangle with the dashed lines represented in Figure 9.

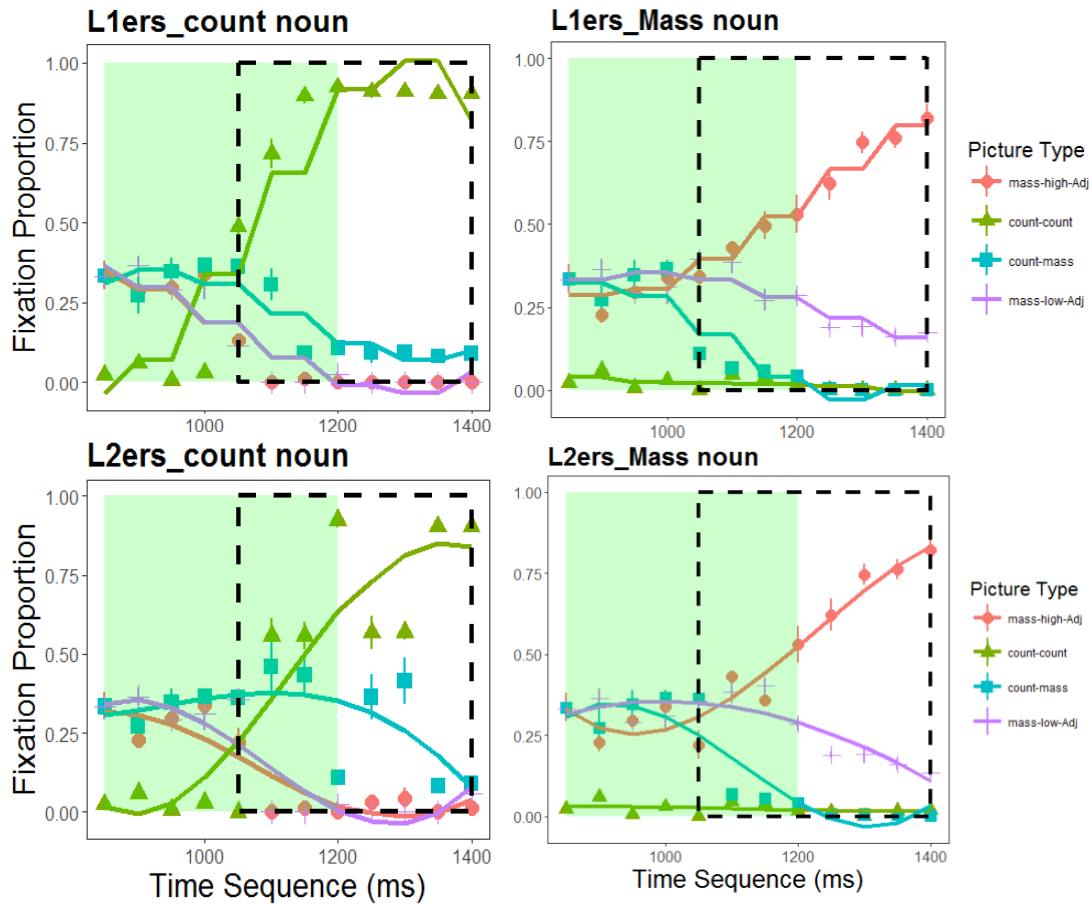


Figure 9. Analysis window of the noun in Condition 3 with gen

For L1ers, when the noun was a count noun, the effect of the picture type on the fixation proportion was significant. There were significantly more fixations in the count-count picture than the mass-high-Adj picture:  $Estimate = 0.412$ ,  $SE = 0.018$ ,  $p < 0.001$ . Also, more fixations were landed in the count-mass picture than the mass-high-Adj picture:  $Estimate = 0.084$ ,  $SE = 0.018$ ,  $p < 0.001$ . The difference of fixation proportions between the mass-high-Adj picture and the mass-low-Adj picture was not significant:  $Estimate = 0.004$ ,  $SE = 0.018$ ,  $p = 0.822$ . When the noun was a mass noun, the effect of the picture type on the fixation proportion was significant. There were significantly less fixations in the count-count picture than the mass-high-Adj picture:  $Estimate = -0.477$ ,  $SE = 0.012$ ,  $p < 0.001$ . Also, less fixations were landed in the count-mass picture than the mass-high-Adj picture:  $Estimate = -0.362$ ,  $SE = 0.012$ ,  $p < 0.001$ . In addition, less fixations were landed in the mass-low-Adj picture than the mass-high-Adj picture:  $Estimate = -0.216$ ,  $SE = 0.012$ ,  $p < 0.001$ . L2ers exhibited the similar patterns to L1ers. When the noun was a count noun, the effect of the picture type on the fixation proportion was significant. There were significantly more fixations in the count-count picture than the mass-high-

Adj picture:  $Estimate = 0.378$ ,  $SE = 0.021$ ,  $p < 0.001$ . Also, more fixations were landed in the count-mass picture than the mass-high-Adj picture:  $Estimate = 0.215$ ,  $SE = 0.019$ ,  $p < 0.001$ . The difference of fixation proportions between the mass-high-Adj picture and the mass-low-Adj picture was not significant:  $Estimate = -0.003$ ,  $SE = 0.018$ ,  $p = 0.914$ . When the noun was a mass noun, the effect of the picture type on the fixation proportion was significant. There were significantly less fixations in the count-count picture than the mass-high-Adj picture:  $Estimate = -0.502$ ,  $SE = 0.018$ ,  $p < 0.001$ . Also, less fixations were landed in the count-mass picture than the mass-high-Adj picture:  $Estimate = -0.491$ ,  $SE = 0.017$ ,  $p < 0.001$ . Besides, less fixations were landed in the mass-low-Adj picture than the mass-high-Adj picture:  $Estimate = -0.235$ ,  $SE = 0.021$ ,  $p < 0.001$ .

These findings illustrate that L1ers and L2ers had the same behaviour patterns in the interest window of the noun: the onset of count nouns drove their attention to the count-noun denoting pictures while the onset of mass nouns drove their attention to the mass-noun denoting pictures. However, different from the situations with the classifier *ba*, when the classifier was *gen*, in Condition 3, on hearing count nouns, both L1ers and L2ers fixed on the count-count picture instead of the count-mass picture even the mass-biased structure is predicted to force a mass/plural-expressing meaning. This is consistent with the results of the Growth Curve Models that different nominal phrase structures affected participants' final choice differently with different classifiers. This unexpected finding will be discussed later in Section 6.

### 5.2.2.3. When the classifier was *kuai*

#### 5.2.2.3.1. In Condition 1 with the structure [Num+CL+N]

The interest window for the noun in Condition 1 is from 750 ms to 1100 ms, as the rectangle with the dashed lines represented in Figure 10.

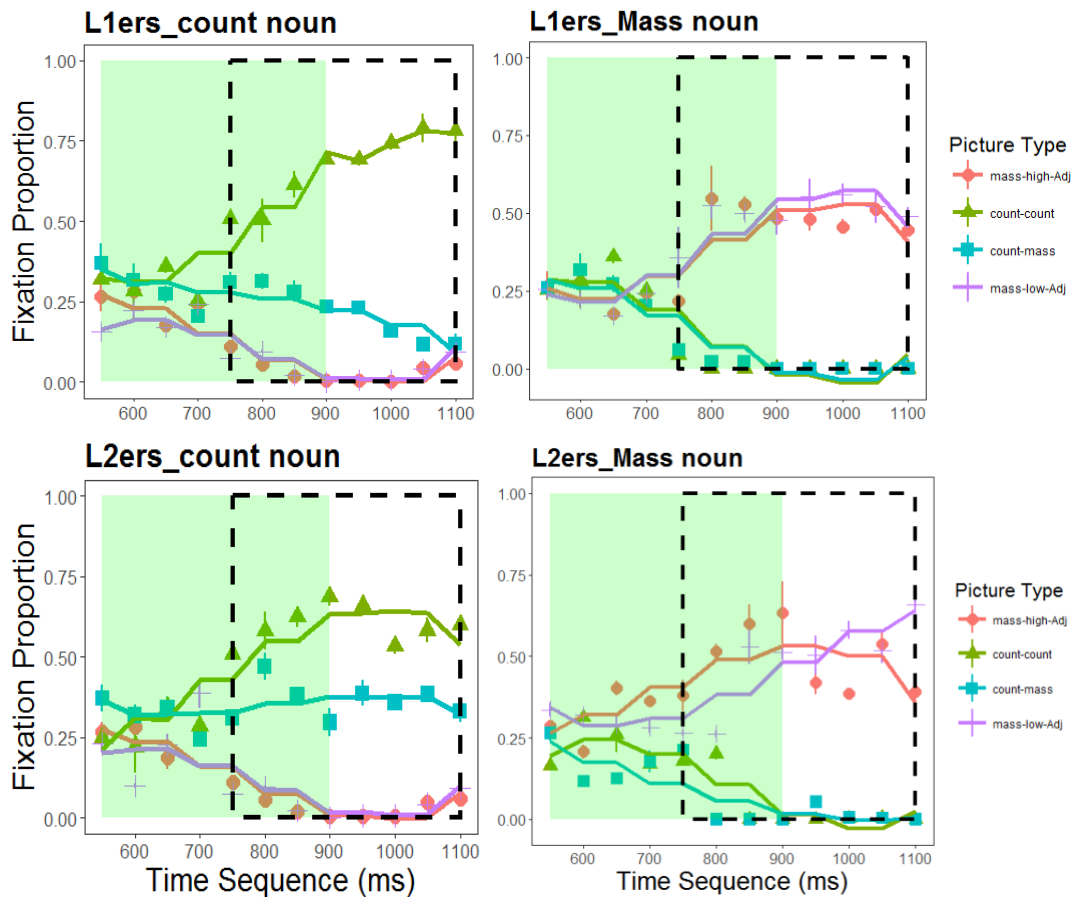


Figure 10. Analysis window of the noun in Condition 1 with kuai

For L1ers, when the noun was a count noun, the effect of the picture type on the fixation proportion was significant. There were significantly more fixations in the count-count picture than in the mass-high-Adj picture:  $Estimate = 0.485$ ,  $SE = 0.014$ ,  $p < 0.001$ . Also, more fixations were landed in the count-mass picture than the mass-high-Adj picture:  $Estimate = 0.201$ ,  $SE = 0.015$ ,  $p < 0.001$ . The difference of fixation proportions between the mass-high-Adj picture and the mass-low-Adj picture was not significant:  $Estimate = -0.004$ ,  $SE = 0.015$ ,  $p = 0.792$ . When the noun was a mass noun, the effect of the picture type on the fixation proportion was significant. There were significantly less fixations in the count-count picture than the mass-high-Adj picture:  $Estimate = -0.301$ ,  $SE = 0.016$ ,  $p < 0.001$ . Also, less fixations were landed in count-mass picture than the mass-high-Adj picture:  $Estimate = -0.382$ ,  $SE = 0.016$ ,  $p < 0.001$ . The difference of fixation proportions between the mass-high-Adj picture and the mass-low-Adj picture was not significant:  $Estimate = 0.015$ ,  $SE = 0.018$ ,  $p = 0.417$ . These observations indicate that count nouns

drove L1ers' attention to the count-noun denoting pictures while mass nouns drove their attention to the mass-noun denoting pictures.

L2ers exhibited the similar patterns with L1ers. When the noun was a count noun, there were significantly more fixations in the count-count picture than in the mass-high-Adj picture:  $Estimate = 0.318$ ,  $SE = 0.019$ ,  $p < 0.001$ . Also, more fixations were landed in the count-mass picture than the mass-high-Adj picture:  $Estimate = 0.243$ ,  $SE = 0.019$ ,  $p < 0.001$ . The difference of fixation proportions between the mass-high-Adj picture and the mass-low-Adj picture was not significant:  $Estimate = 0.004$ ,  $SE = 0.018$ ,  $p = 0.803$ . When the noun was a mass noun, the effect of the picture type on the fixation proportion was significant. There were significantly less fixations in the count-count picture than the mass-high-Adj picture:  $Estimate = -0.238$ ,  $SE = 0.017$ ,  $p < 0.001$ . Also, less fixations were landed in count-mass picture than the mass-high-Adj picture:  $Estimate = -0.245$ ,  $SE = 0.018$ ,  $p < 0.001$ . The difference of fixation proportions between the mass-high-Adj picture and the mass-low-Adj picture was not significant:  $Estimate = 0.017$ ,  $SE = 0.016$ ,  $p = 0.527$ . These observations indicate that similar to L1ers, count nouns drove L2ers' attention to the count noun denoting pictures while mass nouns drove their attention to the mass noun denoting pictures.

#### 5.2.2.3.2. In Condition 2 with the structure [Num+CL+Adj+N]

The interest window for the noun in Condition 2 is from 1050 ms to 1400 ms, as the rectangle with the dashed lines represented in Figure 11.

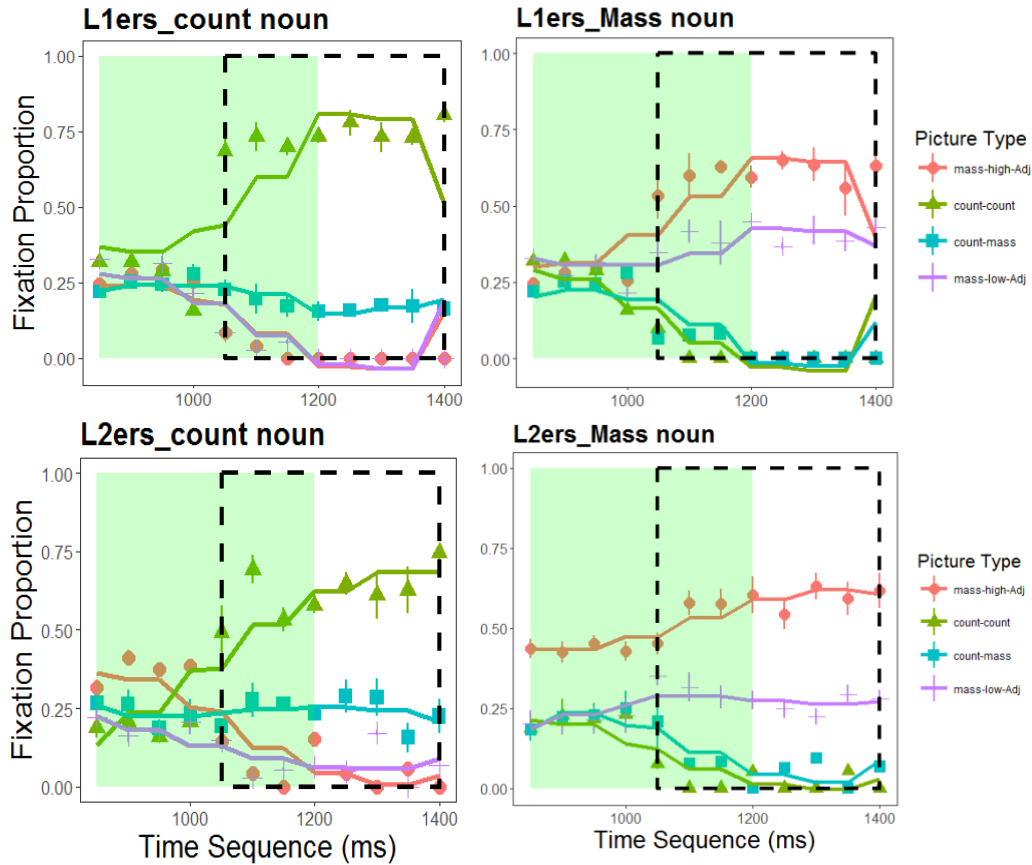


Figure 11. Analysis window of the noun in Condition 2 with kuai

For L1ers, when the noun was a count noun, the effect of the picture type on the fixation proportion was significant. There were significantly more fixations in the count-count picture than the mass-high-Adj picture:  $Estimate = 0.457$ ,  $SE = 0.016$ ,  $p < 0.001$ . Also, more fixations were landed in the count-mass picture than the mass-high-Adj picture:  $Estimate = 0.207$ ,  $SE = 0.016$ ,  $p < 0.001$ . The difference of fixation proportions between the mass-high-Adj picture and the mass-low-Adj picture was not significant:  $Estimate = 0.004$ ,  $SE = 0.016$ ,  $p = 0.872$ . When the noun was a mass noun, the effect of the picture type on the fixation proportion was significant. There were significantly less fixations in the count-count picture than the mass-high-Adj picture:  $Estimate = -0.415$ ,  $SE = 0.017$ ,  $p < 0.001$ . Also, less fixations were landed in the count-mass picture than the mass-high-Adj picture:  $Estimate = -0.398$ ,  $SE = 0.014$ ,  $p < 0.001$ . Besides, less fixations were landed in the mass-low-Adj picture than the mass-high-Adj picture:  $Estimate = -0.111$ ,  $SE = 0.014$ ,  $p < 0.001$ . These observations indicate that the onset of count nouns drove L1ers' attention to the count noun denoting pictures while the onset of mass nouns drove participants' attention to the mass noun denoting pictures.

L2ers exhibited the same patterns with L1ers. When the noun was a count noun, the effect of the picture type on the fixation proportion was significant. There were significantly more fixations in the count-count picture than the mass-high-Adj picture:  $Estimate = 0.367$ ,  $SE = 0.012$ ,  $p < 0.001$ . Also, more fixations were landed in the count-mass picture than the mass-high-Adj picture:  $Estimate = 0.211$ ,  $SE = 0.012$ ,  $p < 0.001$ . The difference of fixation proportions between the mass-high-Adj picture and the mass-low-Adj picture was not significant:  $Estimate = 0.008$ ,  $SE = 0.011$ ,  $p = 0.419$ . When the noun was a mass noun, the effect of the picture type on the fixation proportion was significant. There were significantly less fixations in the count-count picture than the mass-high-Adj picture:  $Estimate = -0.472$ ,  $SE = 0.0017$ ,  $p < 0.001$ . Also, less fixations were landed in the count-mass picture than the mass-high-Adj picture:  $Estimate = -0.413$ ,  $SE = 0.021$ ,  $p < 0.001$ . Less fixations were landed in the mass-low-Adj picture than the mass-high-Adj picture:  $Estimate = -0.174$ ,  $SE = 0.019$ ,  $p < 0.001$ .

#### 5.2.2.3.3. In Condition 3 with the structure [Num+Adj+CL+N]

The interest window for the noun in Condition 3 is from 1050 ms to 1400 ms, as the rectangle with the dashed lines represented in Figure 12.

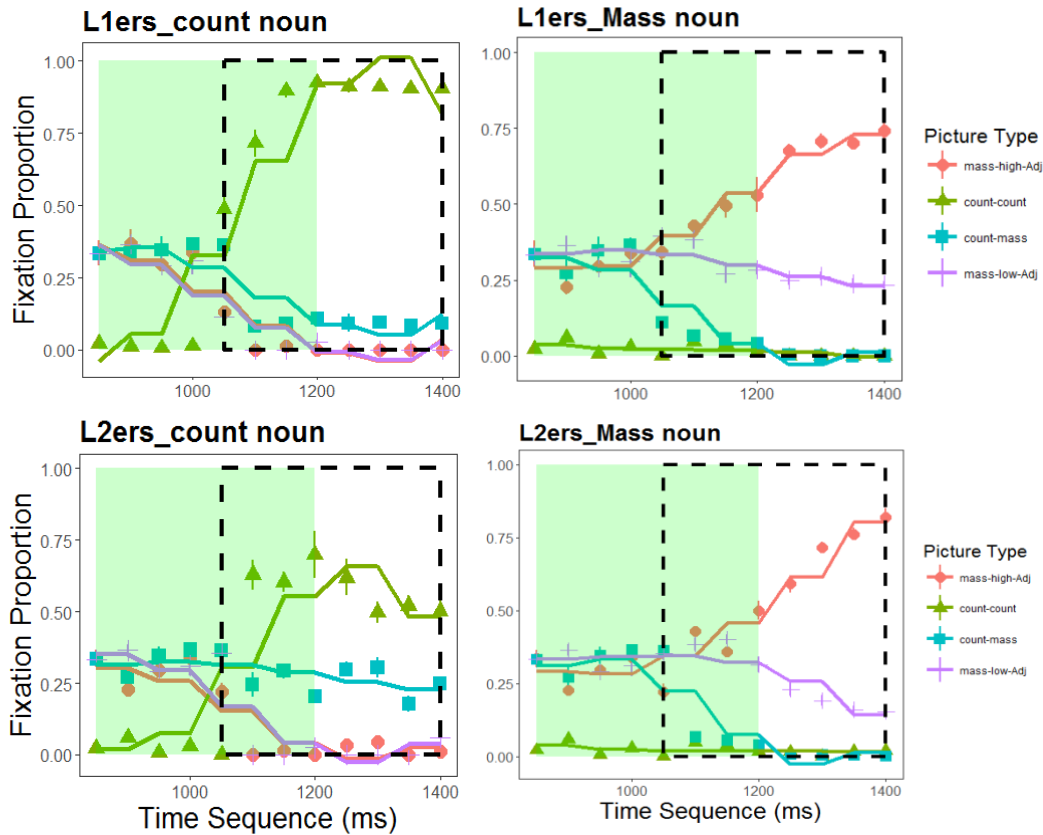


Figure 12. Analysis window of the noun in Condition 3 with kuai



For L1ers, when the noun was a count noun, the effect of the picture type on the fixation proportion was significant. There were significantly more fixations in the count-count picture than the mass-high-Adj picture:  $Estimate = 0.474$ ,  $SE = 0.017$ ,  $p < 0.001$ . Also, more fixations were landed in the count-mass picture than the mass-high-Adj picture:  $Estimate = 0.071$ ,  $SE = 0.018$ ,  $p < 0.001$ . The difference of fixation proportions between the mass-high-Adj picture and the mass-low-Adj picture was not significant:  $Estimate = 0.003$ ,  $SE = 0.017$ ,  $p = 0.848$ . When the noun was a mass noun, the effect of the picture type on the fixation proportion was significant. There were significantly less fixations in the count-count picture than the mass-high-Adj picture:  $Estimate = -0.465$ ,  $SE = 0.017$ ,  $p < 0.001$ . Also, less fixations were landed in the count-mass picture than the mass-high-Adj picture:  $Estimate = -0.425$ ,  $SE = 0.016$ ,  $p < 0.001$ . Besides, less fixations were landed in the mass-low-Adj picture than the mass-high-Adj picture:  $Estimate = -0.228$ ,  $SE = 0.016$ ,  $p < 0.001$ .

L2ers exhibited the similar patterns to L1ers. When the noun was a count noun, the effect of the picture type on the fixation proportion was significant. There were significantly more fixations in the count-count picture than the mass-high-Adj picture:  $Estimate = 0.416$ ,  $SE = 0.021$ ,  $p < 0.001$ . Also, more fixations were landed in the count-mass picture than the mass-high-Adj picture:  $Estimate = 0.278$ ,  $SE = 0.021$ ,  $p < 0.001$ . The difference of fixation proportions between the mass-high-Adj picture and the mass-low-Adj picture was not significant:  $Estimate = -0.005$ ,  $SE = 0.018$ ,  $p = 0.875$ . When the noun was a mass noun, the effect of the picture type on the fixation proportion was significant. There were significantly less fixations in the count-count picture than the mass-high-Adj picture:  $Estimate = -0.478$ ,  $SE = 0.018$ ,  $p < 0.001$ . Also, less fixations were landed in the count-mass picture than the mass-high-Adj picture:  $Estimate = -0.467$ ,  $SE = 0.016$ ,  $p < 0.001$ . Besides, less fixations were landed in the mass-low-Adj picture than the mass-high-Adj picture:  $Estimate = -0.218$ ,  $SE = 0.021$ ,  $p < 0.001$ .

These findings illustrate that L1ers and L2ers had the same behaviour patterns in the interest window of the noun: the onset of count nouns drove their attention to the count noun denoting pictures while the onset of mass nouns drove their attention to the mass noun denoting pictures. However, different from the situations with the classifier *ba* but similar to the situations with the classifier *gen*, when the classifier was *kuai*, in Condition

3, on hearing count nouns, both L1ers and L2ers fixed on the count-count picture instead of the count-mass picture even though the mass-biased structure is predicted to force a mass/plural-expressing meaning. This surprising finding will be discussed later in Section 6.

Taken together, the fixation distributions in the analysis window of the noun are consistent with participants' behavioural results. And also, the division between the classifier *ba* and the classifiers *gen/kuai* is observed in both the on-line data (the fixation distributions in the analysis window of the noun) and the off-line data (participants' behavioural results) on both L1ers and L2ers. It is unclear whether this division between classifiers can also be observed in participants' on-line predictive processing before the appearance of the nouns. To explore this question, both L1ers' and L2ers' fixation distributions in the analysis window of the classifier were analysed.

### **5.3. The division of the classifiers**

Participants' behaviour results and their real time reactions in the analysis window of the noun revealed a significant difference between the classifier *ba* and the classifier *gen* and *kuai*. To further explore different classifiers' effect on participants' predictive processing, the Growth Curve Models were conducted in the interest window of the classifier by separating different classifiers for both L1ers and L2ers.

#### **5.3.1. In Condition 1 with the structure [Num+CL+N]**

In condition 1, the analysis window of the classifier is from 500ms to 750ms, as the rectangle with the dotted lines represented. The fitted curves of proportions of fixations in each picture with different classifiers are plotted in Figure 13.

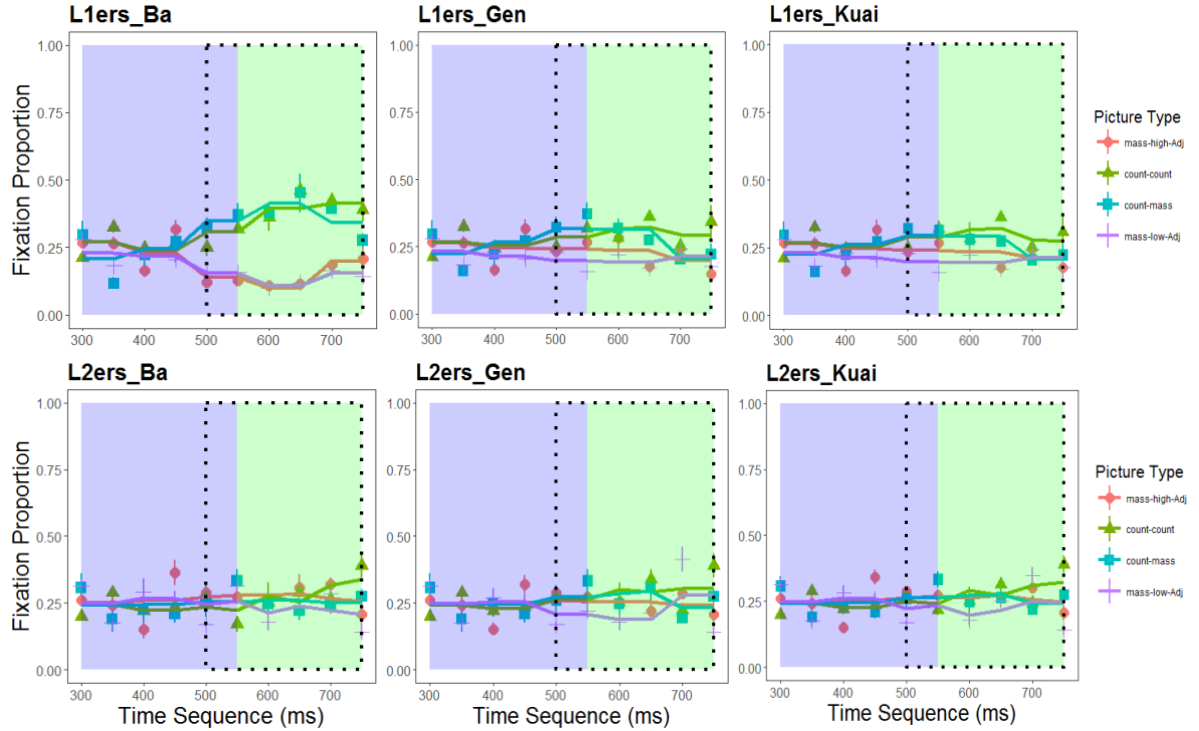


Figure 13. Analysis window of the classifier in Condition 1

From the plots in Figure 13 we can see that for L1ers, the onset of the classifier *ba* triggered different fixation distribution patterns from the classifiers *gen* and *kuai*. When the classifier was *ba*, the GCM revealed a reliable effect of the picture type on the fixation proportions in the analysis window of the classifier for L1ers. There were significantly more fixations in the count-count picture than in the mass-high-Adj picture:  $Estimate = 0.173$ ,  $SE = 0.019$ ,  $p < 0.001$ . Also, more fixations were landed in the count-mass picture than in the mass-high-Adj picture:  $Estimate = 0.171$ ,  $SE = 0.019$ ,  $p < 0.001$ . However, the difference of fixation proportions between the mass-high-Adj picture and the mass-low-Adj picture was not significant:  $Estimate = -0.008$ ,  $SE = 0.019$ ,  $p = 0.649$ . When the classifier was *gen*, there was no significant effect of the picture type on participants' fixation distributions. The difference of fixation proportions between the count-count picture and the mass-high-Adj picture was not significant:  $Estimate = 0.057$ ,  $SE = 0.031$ ,  $p = 0.466$ . The difference between the count-mass picture and the mass-high-Adj picture was not significant:  $Estimate = 0.008$ ,  $SE = 0.031$ ,  $p = 0.785$ . The difference between the mass-high-Adj picture and the mass-low-Adj picture was not significant:  $Estimate = -0.034$ ,  $SE = 0.031$ ,  $p = 0.277$ . When the classifier was *kuai*, similar to the situations when the classifier was *gen*, there was no significant effect of

the picture type on participants' fixation distributions. The difference of fixation proportions between the count-count picture and the mass-high-Adj picture was not significant:  $Estimate = 0.048$ ,  $SE = 0.026$ ,  $p = 0.416$ . The difference of fixation proportions between the count-mass picture and the mass-high-Adj picture was not significant:  $Estimate = 0.012$ ,  $SE = 0.025$ ,  $p = 0.711$ . The difference of fixation proportions between the mass-high-Adj picture and the mass-low-Adj picture was not significant:  $Estimate = -0.042$ ,  $SE = 0.021$ ,  $p = 0.517$ . Comparing the three plots in the top line of Figure 13, we can see a clear pattern of a division between the classifier *ba* and the classifiers *gen* and *kuai*, which has also been observed in participants' behaviour data. These findings indicate that different classifiers affected L1ers' on-line processing and anticipation building. Detailed discussions about this unexpected finding are presented in Section 6.

As for L2ers, different from L1ers, the division of classifier is not observed in the analysis window of the classifier. When the classifier was *ba*, the difference of fixation proportions between the count-count picture and the mass-high-Adj picture was not significant:  $Estimate = 0.041$ ,  $SE = 0.018$ ,  $p = 0.144$ . The difference between the count-mass picture and the mass-high-Adj picture was not significant:  $Estimate = 0.011$ ,  $SE = 0.021$ ,  $p = 0.712$ . The difference between the mass-high-Adj picture and the mass-low-Adj picture was not significant:  $Estimate = -0.025$ ,  $SE = 0.022$ ,  $p = 0.667$ . When the classifier was *gen*, the difference of fixation proportions between the count-count picture and the mass-high-Adj picture was not significant:  $Estimate = 0.042$ ,  $SE = 0.019$ ,  $p = 0.137$ . The difference between the count-mass picture and the mass-high-Adj picture was not significant:  $Estimate = 0.004$ ,  $SE = 0.021$ ,  $p = 0.816$ . The difference between the mass-high-Adj picture and the mass-low-Adj picture was not significant:  $Estimate = -0.045$ ,  $SE = 0.022$ ,  $p = 0.497$ . When the classifier was *kuai*, the difference of fixation proportions between the count-count picture and the mass-high-Adj picture was not significant:  $Estimate = 0.037$ ,  $SE = 0.022$ ,  $p = 0.617$ . The difference of fixation proportions between the count-mass picture and the mass-high-Adj picture was not significant:  $Estimate = 0.008$ ,  $SE = 0.021$ ,  $p = 0.796$ . The difference of fixation proportions between the mass-high-Adj picture and the mass-low-Adj picture was not significant:  $Estimate = -0.006$ ,  $SE = 0.021$ ,  $p = 0.872$ . The three plots in the bottom line of Figure 13 illustrate that different from L1ers, L2ers did not exhibit the division between the classifier *ba* and the classifiers *gen* and *kuai*, which is different from their

behaviour results. Detailed discussions about this surprising finding are presented in Section 6.

### 5.3.2. In Condition 2 with the structure [Num+CL+Adj+N]

In condition 2, the analysis window of the classifier is from 500ms to 750ms, as the rectangle with the dotted lines. The fitted curves of proportions of fixations in each picture with different classifiers are plotted in Figure 14.

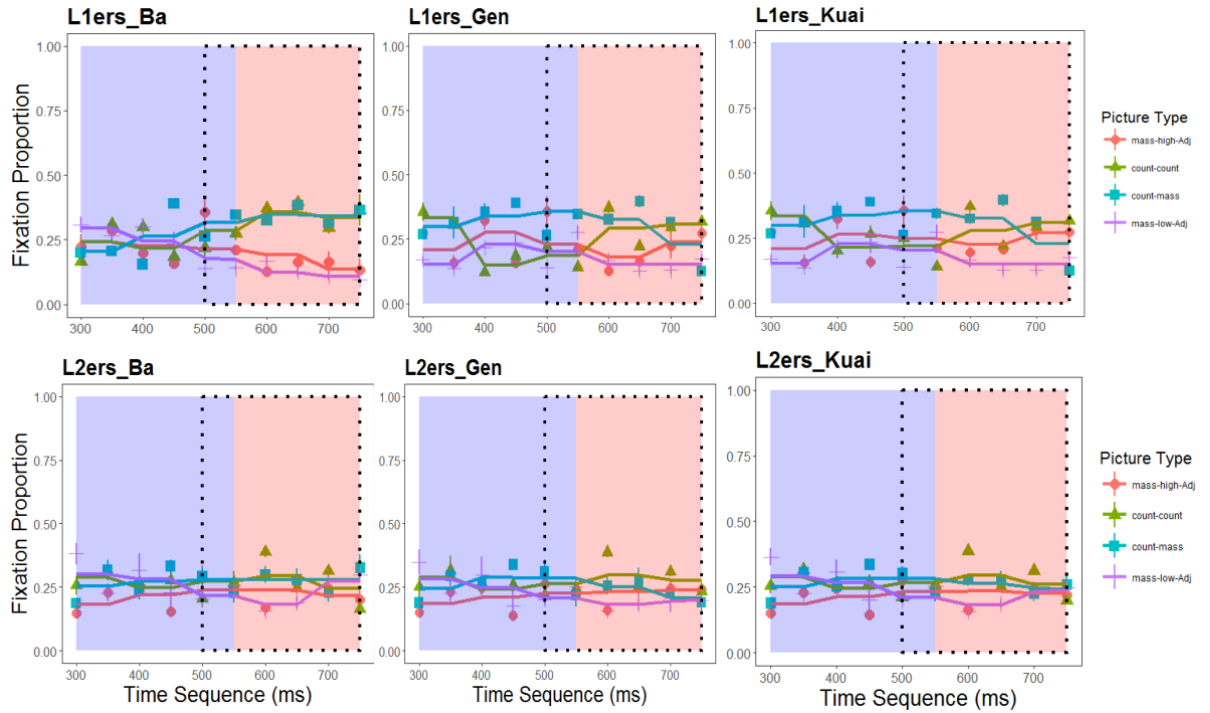


Figure 14. Analysis window of the classifier in Condition 2

From Figure 14 we can see that, similar to the situation in Condition 1, the division of classifiers (*ba* vs. *gen/kuai*) is only observed on L1ers but not L2ers. For L1ers, when the classifier was *ba*, there were significantly more fixations in the count-count picture than in the mass-high-Adj picture: Estimate = 0.085, SE = 0.016,  $p < 0.001$ . Also, more fixations were landed in the count-mass picture than in the mass-high-Adj picture: Estimate = 0.093, SE = 0.016,  $p < 0.001$ . However, the difference of fixation proportions between the mass-high-Adj picture and the mass-low-Adj picture was not significant: Estimate = -0.014, SE = 0.016,  $p = 0.377$ . These results show that, similar to Condition 1, when hearing the classifier *ba*, L1ers landed more fixations in the count-noun denoting pictures than in the mass-noun denoting pictures. When the classifier was *gen*, there was no significant effect of the picture type on the fixation proportions. The difference of

fixation proportions between the count-count picture and the mass-high-Adj picture was not significant:  $Estimate = 0.046$ ,  $SE = 0.018$ ,  $p = 0.173$ . The difference of fixation proportions between the count-mass picture and the mass-high-Adj picture was not significant:  $Estimate = 0.034$ ,  $SE = 0.018$ ,  $p = 0.166$ . The difference of fixation proportions between the mass-high-Adj picture and the mass-low-Adj picture was not significant:  $Estimate = -0.013$ ,  $SE = 0.018$ ,  $p = 0.467$ . When the classifier was *kuai*, there was no significant effect of the picture type on the fixation proportions. The difference of fixation proportions between the count-count picture and the mass-high-Adj picture was not significant:  $Estimate = 0.042$ ,  $SE = 0.018$ ,  $p = 0.185$ . The difference of fixation proportions between the count-mass picture and the mass-high-Adj picture was not significant:  $Estimate = 0.047$ ,  $SE = 0.018$ ,  $p = 0.137$ . The difference of fixation proportions between the mass-high-Adj picture and the mass-low-Adj picture was not significant:  $Estimate = 0.022$ ,  $SE = 0.018$ ,  $p = 0.504$ .

As for L2ers, the influence of the picture type on the fixation proportions was not significant no matter whether the classifier was *ba* or *gen/kuai*. When the classifier was *ba*, the difference of fixation proportions between the count-count picture and the mass-high-Adj picture was not significant:  $Estimate = 0.022$ ,  $SE = 0.017$ ,  $p = 0.335$ . The difference of fixation proportions between the count-mass picture and the mass-high-Adj picture was not significant:  $Estimate = 0.028$ ,  $SE = 0.016$ ,  $p = 0.398$ . The difference of fixation proportions between the mass-high-Adj picture and the mass-low-Adj picture was not significant:  $Estimate = -0.008$ ,  $SE = 0.014$ ,  $p = 0.819$ . When the classifier was *gen*, the difference of fixation proportions between the count-count picture and the mass-high-Adj picture was not significant:  $Estimate = 0.031$ ,  $SE = 0.014$ ,  $p = 0.713$ . The difference of fixation proportions between the count-mass picture and the mass-high-Adj picture was not significant:  $Estimate = 0.028$ ,  $SE = 0.016$ ,  $p = 0.421$ . The difference of fixation proportions between the mass-high-Adj picture and the mass-low-Adj picture was not significant:  $Estimate = -0.007$ ,  $SE = 0.016$ ,  $p = 0.719$ . When the classifier was *kuai*, the difference of fixation proportions between the count-count picture and the mass-high-Adj picture was not significant:  $Estimate = 0.027$ ,  $SE = 0.016$ ,  $p = 0.628$ . The difference of fixation proportions between the count-mass picture and the mass-high-Adj picture was not significant:  $Estimate = 0.021$ ,  $SE = 0.016$ ,  $p = 0.531$ . The difference of fixation proportions between the mass-high-Adj picture and the mass-low-Adj picture was not significant:  $Estimate = 0.008$ ,  $SE = 0.016$ ,  $p = 0.871$ .

### 5.3.3. In Condition 3 with the structure [Num+Adj+CL+N]

In condition 3, the analysis window of the classifier is from 800ms to 1050ms, as the rectangle with the dotted lines. The fitted curves of proportions of fixations in each picture with different classifiers are plotted in Figure 15.

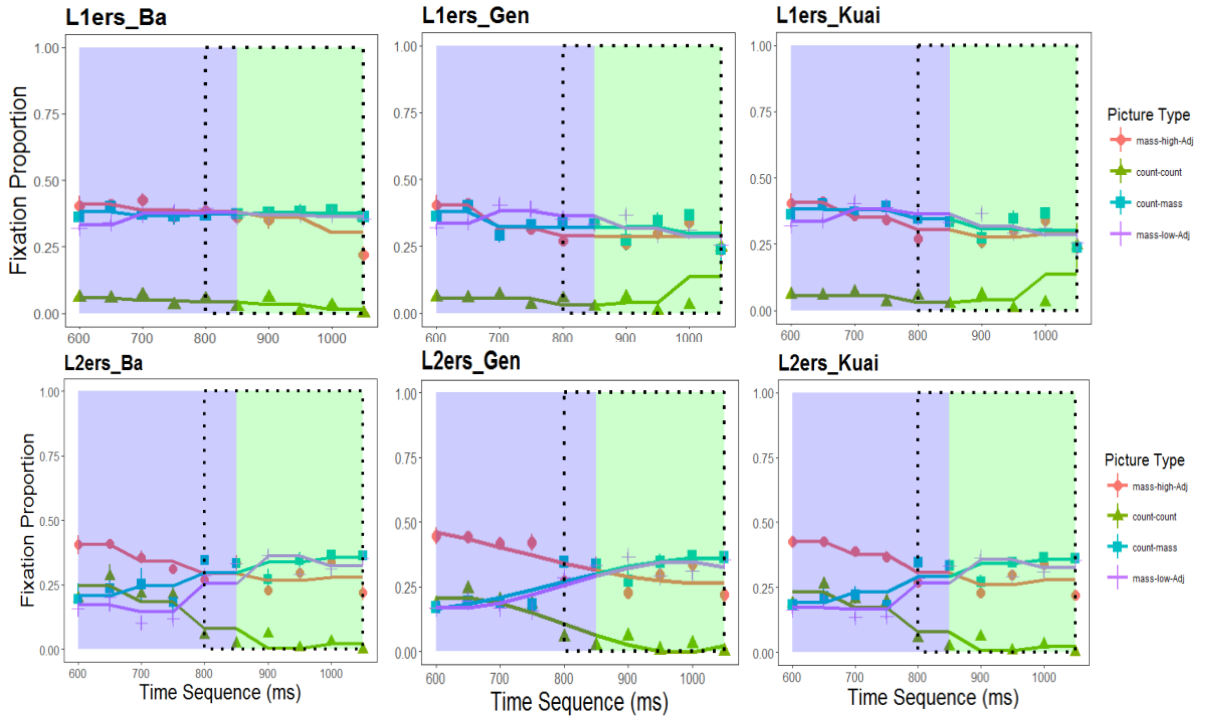


Figure 15. Analysis window of the classifier in Condition 3

From the plots in Figure 15 we can see that, in Condition 3 where the nominal phrases have the structure [Num+Adj+CL+N], the classifiers occurred following the adjectives affected both L1ers and L2ers in the same way across all three classifiers. The division between the classifier *ba* and the classifiers *gen* and *kuai* was not found in Condition 3. In the analysis window of the classifier, both L1ers and L2ers landed more fixations in the mass/plural-expressing pictures (the two mass-noun denoting pictures and the count-mass picture) than the count-expressing picture (the count-count picture). For L1ers, when the classifier was *ba*, there were significantly less fixations in the count-count picture than in the mass-high-Adj picture:  $Estimate = -0.329$ ,  $SE = 0.008$ ,  $p < 0.001$ . The difference of fixation proportions between the count-mass picture and the mass-high-Adj picture was not significant:  $Estimate = 0.006$ ,  $SE = 0.008$ ,  $p = 0.416$ . Also, the difference of fixation proportions between the mass-high-Adj picture and the mass-low-Adj picture was not significant:  $Estimate = -0.004$ ,  $SE = 0.008$ ,  $p = 0.573$ . This result indicates that

the classifiers in Condition 3 did not affect L1ers' interpretation of the nominal phrases. The onset of the adjectives already helped them to build a mass-biased structure in which only massifiers can fit in, and on hearing the classifiers, they automatically interpreted the classifiers as massifiers and expected the nominal phrase to have a mass/plural-expressing meaning. Consequently, their fixations focused on the three mass/plural-expressing pictures since the onset of the adjectives, and kept the same pattern in the interest window of the classifier. When the classifier was *gen*, there were significantly less fixations in the count-count picture than the mass-high-Adj picture:  $Estimate = -0.254$ ,  $SE = 0.012$ ,  $p < 0.001$ . The difference of fixation proportions between the count-mass picture and the mass-high-Adj picture was not significant:  $Estimate = 0.012$ ,  $SE = 0.012$ ,  $p = 0.349$ . Also, the difference of fixation proportions between the mass-high-Adj picture and the mass-low-Adj picture was not significant:  $Estimate = 0.019$ ,  $SE = 0.012$ ,  $p = 0.125$ . When the classifier was *kuai*, there were significantly less fixations in the count-count picture than the mass-high-Adj picture:  $Estimate = -0.304$ ,  $SE = 0.015$ ,  $p < 0.001$ . The difference of fixation proportions between the count-mass picture and the mass-high-Adj picture was not significant:  $Estimate = 0.009$ ,  $SE = 0.014$ ,  $p = 0.591$ . Also, the difference of fixation proportions between the mass-high-Adj picture and the mass-low-Adj picture was not significant:  $Estimate = 0.011$ ,  $SE = 0.012$ ,  $p = 0.359$ .

L2ers exhibited the native-like pattern in the analysis window of the classifier with all three classifiers. When the classifier was *ba*, there were significantly less fixations in the count-count picture than in the mass-high-Adj picture:  $Estimate = -0.308$ ,  $SE = 0.011$ ,  $p < 0.001$ . The difference of fixation proportions between the count-mass picture and the mass-high-Adj picture was not significant:  $Estimate = 0.012$ ,  $SE = 0.013$ ,  $p = 0.539$ . Also, the difference of fixation proportions between the mass-high-Adj picture and the mass-low-Adj picture was not significant:  $Estimate = -0.007$ ,  $SE = 0.008$ ,  $p = 0.547$ . When the classifier was *gen*, significantly less fixations were landed in the count-count picture than the mass-high-Adj picture:  $Estimate = -0.371$ ,  $SE = 0.016$ ,  $p < 0.001$ . The difference of fixation proportions between the count-mass picture and the mass-high-Adj picture was not significant:  $Estimate = 0.018$ ,  $SE = 0.014$ ,  $p = 0.413$ . Also, the difference of fixation proportions between the mass-high-Adj picture and the mass-low-Adj picture was not significant:  $Estimate = 0.018$ ,  $SE = 0.013$ ,  $p = 0.421$ . When the classifier was *kuai*, significantly less fixations were landed in the count-count picture than the mass-high-Adj picture:  $Estimate = -0.318$ ,  $SE = 0.016$ ,  $p < 0.001$ . The difference of fixation



proportions between the count-mass picture and the mass-high-Adj picture was not significant:  $Estimate = -0.008$ ,  $SE = 0.015$ ,  $p = 0.476$ . Also, the difference of fixation proportions between the mass-high-Adj picture and the mass-low-Adj picture was not significant:  $Estimate = -0.009$ ,  $SE = 0.016$ ,  $p = 0.471$ .

In Condition 3, in the interest window of the classifier, L2ers exhibited the same patterns as L1ers', indicating that similar to L1ers, when realizing the structure is a mass-biased structure (on hearing the Adj-CL word order), L2ers expected the nominal phrase to have mass/plural-expressing meanings, and consequently, fixed on the three mass/plural-expressing pictures. This native-like behaviour in the interest window of the classifier for L2ers illustrates that they were sensitive to the Adj-CL word order and can use this mass-preferred syntactic cue to build anticipations in real time processing.

In general, the analysis of participants' fixation distributions in the analysis window of the classifier indicates that different classifiers affected L1ers' predictive processing: on hearing *ba*, they started to fix more on the count-noun denoting pictures than the mass-noun denoting pictures in the neutral structures (Condition 1 and Condition 2), while the onset of the classifier *gen/kuai* did not trigger any preference for any picture. This is consistent with their behaviour results that L1ers reacted differently to nominal phrase structures with different classifiers. As for L2ers, even though they exhibited the division between *ba* and *gen/kuai* in their final picture choices, this classifier division effect is not found in their on-line processing in the analysis window of the classifier. Detailed discussions are presented in Section 6.

#### 5.3.4. Further explorations of the division of classifiers

It is possible that the difference between the classifier *ba* and the other two classifiers *gen/kuai* is caused by their different types of ambiguity. The three dual-role classifiers used in the current study are all ambiguous words since they allow for both count-classifier readings and massifier readings. According to Rayner & Duffy (1986), there are two kinds of ambiguous words: those for which all the possible interpretations of the ambiguous word are equally likely, and those for which there is one highly dominant interpretation among all the possible options.

Since these three dual-role classifiers could have either count-classifier meanings or massifier meanings, there are three possibilities about their ambiguity types: count-

biased classifiers with the dominant count-classifier meanings, mass-biased classifiers with the dominant massifier meanings, and equally-biased classifiers which have equally likely count-classifier meanings and massifier meanings. To inspect what ambiguity type each classifier is, it is necessary to know what is the preferred interpretation of each dual-role classifier by native Mandarin speakers. Apart from explicitly asking native Mandarin speakers what is the preferred interpretation of a specific classifier, which is not a very reasonable test since the information about count/mass classifiers is implicit to Mandarin natives (as discussed in Chapter 2 Section 3, they use it unconsciously and automatically in daily express but without any explicit and conscious awareness about what is a count/mass classifier), the ambiguity type of each dual-role classifier can be reflected by native Mandarin speakers' interpretations of nominal phrases. For example, if most Mandarin natives interpret the nominal phrase *yi ba yaoshi* 'one CI key' as 'one key', it means the preferred interpretation of the classifier *ba* is the count-classifier meaning; on the contrary, if most of them interpret it as 'one handful of keys', then the preferred interpretation of the classifier *ba* is the massifier meaning. Also, apart from the preferred interpretations of nominal phrases, the information of what category (mass/count) of nouns is preferred to occur following each classifier can also help us to gain some understandings of classifiers' ambiguity type.

Chou, Huang, Lee, & Lee (2014) used a Cloze Probability Test to examine the cloze probability of nouns following specific classifiers. In their study, participants were asked to give three completions to the fragment of 'numeral+classifier+( )'. The authors argued that the first noun participants filled in is the first noun coming to their mind when reading the classifier, which can be treated as the best completion. And the second noun is the second best completion, and so on. The cloze probability was calculated for both the best completion and the second best completion. For example, among 30 participants, 25 of them filled the noun *yaoshi* 'key' as the best completion after the classifier *ba*, 3 of them filled *yaoshi* 'key' as the second best completion, the cloze probability of *yaoshi* 'key' being the best and the second best completion are 83.33% (25/30) and 10% (3/30), respectively. And the overall cloze probability for the noun *yaoshi* 'key' is 93.33% (83.33% + 10%).

Following Chou et al. (2014), in order to gain the information about the ambiguity types of the classifiers and their preference of mass/count category of nouns, a revised Cloze Probability Test was run on 40 native Mandarin speakers (none of them had participated

in the two eye-tracking experiments, the Noun Rating Test, and the Norming Test), in which they were asked to fill in the blank in the construction [yi ‘one’+CL+( )], using at least five nouns. After each noun, they were also required to suggest how many items the phrase refers to by writing down a corresponding number, or a question mark when they were not sure about the exact number of the items. For example, when the given construction is *yi ba* ( ), a possible noun which can combine with *ba* is *yaoshi* ‘key’. If participants interpret the phrase *yi ba yaoshi* as ‘one key’ (in which the classifier *ba* is interpreted as a count-classifier), they should write the number 1 after the noun *yaoshi* ‘key’. However, if they interpret the phrase as ‘one handful of keys’ (in which the classifier *ba* is interpreted as a massifier), they should write a question marker after the noun since it is unclear how many keys there are in a handful of keys. Through this revised Cloze Probability Test we can gain two kinds of information: what category of nouns (count vs. mass) native Mandarin speakers prefer to use after each classifier (the mass/count status of the filled nouns are judged based on the results of the Noun Rating Test), and what is the dominant meaning (count-classifier vs. massifier) of the classifier.

The criteria for count-biased classifier is different from that of mass-biased and equally-biased classifiers. A classifier is a count-biased classifier if most native Mandarin speakers filled in count nouns after it and interpreted these nominal phrases as referring to one individual object by writing down the numeral 1 after the nouns. To be specific, a classifier is a count-biased classifier if 1). the overall cloze probability of count nouns is more than 70%, and 2). the numeral 1 is labelled following more than 70% of the count nouns. The rationale is if most native Mandarin speakers fill in count nouns after a classifier, it means the first several nouns coming to their mind on reading the classifier are count nouns, indicating that this classifier has a strong connection with count nouns. Further, if most nominal phrases are interpreted as referring to one individual object reflecting by the numeral 1 after the nouns, it means that the dominant meaning of the dual-role classifier is the count-classifier meaning which describes the solid individual unit objects naturally have. However, this criterion cannot be used for the mass-biased and equally-biased classifiers. As mentioned in Chapter 2, in Mandarin Chinese, count-classifiers occur with only count nouns while massifier occur with either count nouns or mass nouns. Thus the mass/count status of the best completions cannot be direct evidence of the ambiguity type of mass-/equally-biased classifiers. Also, neither the number 1 nor the question marker following the nouns can offer us useful information about the

dominant meanings of mass-/equally-biased classifiers. For example, in the nominal phrase *yi gen qianbi* ‘one pencil’, the noun *qianbi* ‘pencil’ is a count noun and the classifier *gen* is a count-classifier since it describes the natural unit a pencil naturally possesses. This nominal phrase *yi gen pianbi* ‘one pencil’ refers to one single pencil thus can be labelled by the numeral 1. In the nominal phrase *yi gen toufa* ‘one strand of hair’, the noun *toufa* ‘hair’ is a mass noun and the classifier *gen* is a massifier since it creates a measurement – ‘strand’ – to make the substance ‘hair’ countable. This nominal phrase *yi gen toufa* ‘one strand of hair’ refers to one single piece of hair thus can also be labelled by the numeral 1. Thus no matter whether *gen* is interpreted as a count-classifier or a massifier, the numeral 1 can be labelled following the nouns. Hence, unfortunately, from this revised Cloze Probability Test, it is not clear whether a classifier is a mass-biased classifier or an equally-biased classifier. The only thing can be assured from this revised Cloze Probability Test is whether a classifier is a count-biased classifier or not.

An example of the revised Cloze Probability Test and participants’ answers is illustrated in (38).

(38) a. *Yi ba (jita, 1; jiandao, 1; yusan, 1; shuzi, 1; huo, ?; yaoshi, 1; chaopiao, ?)*

One CL\_ba (guitar, 1; scissors, 1; umbrella, 1; comb, 1; fire, ?; key, 1; cash, ?)

b. *Yi gen (shengzi, 1; qianbi, 1; xuegao, 1; toufa, 1; huochai, 1)*

One CL\_rod (string, 1; pencil, 1; popsicle, 1; hair, 1; match, 1)

c. *Yi kuai (shoujuan, 1; qiaokeli, ?; boli, 1; wudian, ?; dangao, ?)*

One CL\_chunk (handkerchief, 1; chocolate, ?; glass, 1; stain, ?; cake, ?)

Since we only care about what category of nouns (count vs. mass) native Mandarin speakers preferred to fill in after each classifier, instead of the cloze probability of each specific noun, the cloze probability of count nouns and mass nouns were calculated separately. Among the five nouns, the cloze probability of the best, the second best, the third best completions, and the overall cloze probability (the average of the first three best completions) for both count nouns and mass nouns after each classifier are summarized in Table 16.

Table 16. The cloze probability of count and mass nouns

Classifier	Count nouns				Mass nouns			
	1 <sup>st</sup> (%)	2 <sup>nd</sup> (%)	3 <sup>rd</sup> (%)	Overall	1 <sup>st</sup> (%)	2 <sup>nd</sup>	3 <sup>rd</sup> (%)	Overall
<i>Ba</i>	95	87.5	75	<b>85.83</b>	5	12.5	25	14.16
<i>Gen</i>	55	40	50	48.4	45	60	50	51.6
<i>Kuai</i>	47.5	55	45	49.17	52.5	45	55	50.83

From Table 16 we can see that the classifier *ba* was more preferred to be followed by count nouns than mass nouns (85.83% vs. 14.16%), while the classifiers *gen/kuai* did not have any preference for either count nouns or mass nouns (48.4% vs. 51.6%, 49.17% vs. 50.83%, respectively). This finding indicates that *ba* evokes a preference for count-category nouns on participants, while *gen/kuai* are happy to occur with either count nouns or mass nouns. Even though more count nouns are preferred to occur with the classifier *ba* than mass nouns, one cannot reach the conclusion that *ba* is a count-biased classifier. Since even followed by count nouns, native Mandarin speakers may still interpret it as a massifier. For example, in the nominal phrase *yi ba qianbi* with the interpretation of ‘one handful of pencils’, *pianbi* ‘pencil’ is a count noun but *ba* is a massifier. To figure out what is the dominant interpretation of the classifier *ba*, the cloze probability of count/mass nouns and the actual interpretations of the nominal phrases (the percentage of the numeral 1 and the question mark following each noun, illustrated in Table 17) need to be analysed together.

Table 17. Interpretation of nominal phrases (%)

	<i>Ba</i>		<i>Gen</i>		<i>Kuai</i>	
	Count Noun	Mass Noun	Count Noun	Mass Noun	Count Noun	Mass Noun
<b>1</b>	99.16	0.79	98.77	90.53	99.49	87.69
<b>?</b>	0.84	99.21	1.23	9.47	0.51	12.31

From Table 16 we can see that the overall cloze probability of count nouns following the classifier *ba* is 85.33%. And from Table 17 we can see that 99.16% of the count nouns following *ba* were followed by the numeral 1. Based on these results I argue that the classifier *ba* is a count-biased classifier: it is more preferred to be interpreted as a count-

classifier than a massifier, and has a strong ‘priming effect’ for its associated count nouns since on reading it, associated count nouns occurred in native speakers’ mind earlier and faster than possible associated mass nouns.

To sum up, the results of the revised Cloze Probability Test indicate that, among the three dual-role classifiers used in the current study (*ba*, *gen*, and *kuai*), *ba* is a count-biased classifier while the classifiers *gen* and *kuai* are non-count-biased classifiers (it is not clear whether they are mass-biased classifiers or equally-biased classifiers).

## 6. Discussion

In the current experiment, the Visual World Paradigm was used to look into native Mandarin speakers’ (L1ers) and high proficiency late Dutch-Mandarin learners’ (L2ers) predictive using of the mass/count syntactic cues. Three dual-role classifiers were used: *ba* ‘handful’, *gen* ‘rod’, and *kuai* ‘chunk’. There were 24 nouns which could occur with one of the three dual-role classifiers, half of them are typical count nouns and the other half are typical mass nouns based on the results from the Noun Rating test. For each nominal phrase, there were three conditions of the nominal phrase structures: nominal phrases in Condition 1 have the structure [Num+CL+N], in Condition 2 have the structure [Num+CL+Adj+N], and in Condition 3 have the structure [Num+Adj+CL+N]. By comparing Condition 1 to Condition 2, one can tell whether adding an adjective which modifies the noun would affect L1ers’ and L2ers’ processing of nominal phrases. By comparing Condition 2 to Condition 3 one can tell whether different nominal phrase structures (CL-Adj vs. Adj-CL) would affect participants’ predictive processing and interpretation.

For each pair of nouns, there were four pictures presented on the screen. Two of them are count-noun denoting pictures while the other two are mass-noun denoting pictures. By manipulating the properties of the pictures and the nominal phrase structures in the audio materials, an underlying probability distributions of the locations of participants’ fixations were created between the former and the latter. By observing the change of participants’ fixation proportions among the four pictures along time, one can tell how the onset of each item in the nominal phrase affected participants’ on-line processing.

In general, there are two parts of results in the Visual World Paradigm experiment: behaviour results and fixation distributions. Behaviour results reflect participants' final interpretations of nominal phrases. Fixation distributions reveal the influence of each item in nominal phrases on participants' predictive processing.

As reported in Section 5 that both behaviour results and fixation distributions uncover a division between the classifier *ba* and the classifiers *gen/kuai*. In the following part of this section, this classifier division effect is discussed first, followed by the discussions about participants' behaviour results and fixation distributions.

### 6.1. The division of the classifiers

The results in the Visual World Paradigm experiment illustrate a clear division between the classifier *ba* and the classifiers *gen/kuai* in both L1ers' and L2ers' final picture choices and in L1ers' real time fixation distributions. The behavioural results show that, when nominal phrases have the structure [Num+Cl+Adj+N] (in Condition 2) and with mass nouns, both L1ers and L2ers chose the mass-low-Adj picture as the target picture when the classifier was *ba*, while they chose the mass-high-Adj picture when the classifier was *gen/kuai*. When nominal phrases have the structure [Num+Adj+Cl+N] (in Condition 3) and with count nouns, both L1ers and L2ers chose the count-mass picture as the target picture when the classifier was *ba*, while choosing the count-count picture when the classifier was *gen/kuai*. The fixation distributions in the analysis window of the classifier show that when the nominal phrases have the structure [Num+Cl+N] (in Condition 1) or the structure [Num+Cl+Adj+N] (in Condition 2), on hearing the classifier *ba* L1ers started to direct their attention to the count-noun denoting pictures, while the onset of the classifier *gen/kuai* did not exhibit this preference for count-noun denoting pictures. These observations indicate that *ba* is different from *gen* and *kuai*, and this difference between classifiers has been observed consistently in both behavioural data and real time predictive processing results. Before more discussions about the two main parts of the results (i.e. behavioural results and fixation distributions), it is necessary to figure out the reason of the particularity of the classifier *ba*.

Firstly, it should be noted that although the three classifiers are all dual-role classifiers, they have different functions when being massifiers: being a massifier, *ba* behaves like a collective classifier which measures objects by grouping them together, while *gen* and *kuai* behave like dividers which measure objects by dividing them apart (Cheng, 2010).

For example, in the nominal phrase *yi ba qianbi* ‘one handful of pencils’, the classifier *ba* is doing a massifier’s job by measuring multiple pencils in handfuls. In the nominal phrase *yi kuai nailao* ‘one chunk of cheese’, the classifier *kuai* is doing a massifier’s job of measuring cheese by dividing it to chunks. This difference between the collective classifier *ba* and the dividers *gen/kuai* may cause the difference in participants’ final choice and their fixation distributions.

Secondly, the fixation distribution differences between the classifier *ba* and the other two classifiers *gen/kuai* may be caused by their different types of ambiguity. Based on the Cloze Probability Test reported in Section 5, the classifier *ba* is a count-biased classifier, while the classifiers *gen* and *kuai* could be either mass-biased classifiers or equally-biased classifiers. These two groups of classifiers differ in their preference of occurring with count nouns or mass nouns, and native speakers’ preference of interpreting them as count-classifiers or massifiers.

Apart from the massifier meanings (collective vs. divider), the ambiguity types (count-/mass-/equally-biased), and preferred count/mass nouns, the three classifiers also differ in their co-occurrence frequency with count/mass nouns. Using two Mandarin corpora (National Language Resources Monitoring and Research Centre, Broadcast Media Language Branch, 2009; Ministry of Education and institute of Applied Linguistics, 2009) and the number of Google hits (February 2017) (Blair, Urland, & Ma, 2002; Pollatsek, Drieghe, Stockall, & De Almeida, 2010), the co-occurrence frequency counts of the count and mass nouns used in the current research with the three classifiers were calculated. The averages of the co-occurrence frequency of count nouns and mass nouns following these three classifiers are illustrated in Table 18.

Table 18. The averages of co-occurrence frequency (per million)

	<i>Ba</i>			<i>Gen</i>			<i>Kuai</i>		
	Count	Mass	<i>diff</i>	Count	Mass	<i>diff</i>	Count	Mass	<i>diff</i>
<b>Frequency</b>	0.689	0.641	-	0.356	0.513	0.157*	0.432	0.594	0.162*

\*  $p < 0.01$

The results show that when the classifier is *ba*, the difference of the co-occurrence frequency between count nouns and mass nouns is not significant,  $p = 0.347$ ; while when



the classifier is *gen/kuai*, mass nouns have significantly higher co-occurrence frequency than count nouns,  $ps < 0.01$ .

Taken together, the differences between the classifier *ba* and the classifiers *gen/kuai* are illustrated in Table 19.

Table 19. The differences between *ba* and *gen/kuai*

	Massifier meaning	Ambiguity type	Preferred nouns	Co-occurrence frequency
<b><i>Ba</i></b>	Collective	Count-biased	Count nouns	Count = Mass
<b><i>Gen</i></b>	Divider	Mass/Equal-biased	Count/Mass nouns	Count < Mass
<b><i>Kuai</i></b>	Divider	Mass/Equal-biased	Count/Mass nouns	Count < Mass

From Table 19 we can see that the division of the three dual-role classifiers based on participants' behavioural results and fixation distributions is consistent with the division based on their massifier meanings, ambiguity types, preferred nouns' count/mass categories, and the co-occurrence frequency with count and mass nouns. This finding indicates that apart from the nominal phrase structure which is a manipulated factor in the current study, the properties of different classifiers (e.g. ambiguity types) and the restrictions of classifier-noun pairings (e.g. classifiers' preferred nouns, and the co-occurrence frequency of mass/count nouns following each classifier) also have some effects on participants' real time processing and anticipation building. The observed division between classifiers is caused by their different properties.

In the following part of this section, participants' behaviour results and fixation distributions are discussed based on the influence of mass/count structures as well as the properties of classifiers.

## 6.2. Behaviour results

Participants' behaviour data is summarized in Table 15 and repeated here as in Table 20. From the comparison between L1ers' and L2ers' we can see that, the two groups exhibited the same behaviour patterns. Also, both of them exhibited the classifier division effect: the pattern with the classifier *ba* was different from the pattern with the classifier *gen* and *kuai*.

Table 20. L1ers' and L2ers' behaviour results

<i>ba</i>	Native Mandarin speakers						Late Dutch-Mandarin learners					
	Condition 1		Condition 2		Condition 3		Condition 1		Condition 2		Condition 3	
	Count	Mass	Count	Mass	Count	Mass	Count	Mass	Count	Mass	Count	Mass
<b>M-hi</b>	0 %	<b>51%</b>	0 %	25%	0 %	<b>95%</b>	0%	<b>55%</b>	0%	40%	0%	<b>80%</b>
<b>C-C</b>	<b>80%</b>	0 %	<b>95%</b>	0%	1%	0 %	<b>84%</b>	0%	<b>80%</b>	0%	5%	0%
<b>C-M</b>	20%	0 %	5%	0 %	<b>99%</b>	0 %	16%	0%	20%	0%	<b>95%</b>	0%
<b>M-lo</b>	0%	49%	0 %	<b>75%</b>	0 %	5%	0%	45%	0%	<b>60%</b>	0%	20%
<i>Gen</i>	Condition 1		Condition 2		Condition 3		Condition 1		Condition 2		Condition 3	
<i>kuai</i>	Count	Mass	Count	Mass	Count	Mass	Count	Mass	Count	Mass	Count	Mass
<b>M-hi</b>	0 %	<b>55%</b>	0 %	<b>75%</b>	0 %	<b>80%</b>	0%	50%	0%	<b>80%</b>	0%	<b>60%</b>
<b>C-C</b>	<b>90%</b>	0 %	<b>90%</b>	0 %	<b>95%</b>	0 %	<b>95%</b>	0%	<b>95%</b>	0%	<b>95%</b>	0%
<b>C-M</b>	10%	0 %	10%	0 %	5%	0 %	5%	0%	5%	0%	5%	0%
<b>M-lo</b>	0%	45%	0 %	25%	0 %	20%	0%	50%	0%	20%	0%	40%

(*M-hi* stands for the mass-high-Adj picture, *M-lo* stands for the mass-low-Adj picture, *C-C* stands for the count-count picture, *C-M* stands for the count-mass picture)

From Table 20 we can see that, when the classifier was *ba*, nominal phrases with different structures (mass/count syntactic cues) affected participants' interpretations. When the noun was a count noun, in both Condition 1 and Condition 2, participants preferred the count-count picture as the target picture, while in Condition 3, participants preferred the count-mass picture. Both of the two pictures contain objects denoted by the same count nouns associated with *ba*. The difference between these two pictures is how the objects are organized: in the count-count picture, the objects are presented in their natural units, while in the count-mass picture, the objects are organized in groups. By choosing the count-count picture as the target picture, participants interpreted the

nominal phrase as expressing countable objects with discrete units, while choosing the count-mass picture, participants interpreted the nominal phrases as expressing quantitative meanings, in other words, objects organized in groups and counted by the groups instead of their natural units. Thus when the nominal phrases have the structure [Num+CL+N] or [Num+CL+Adj+N] (i.e. there is no mass syntactic cue), both L1ers and L2ers preferred to interpret them as countable objects with discrete units which these objects naturally have, and chose the count-count picture to be the target. However, when the nominal phrases have the structure [Num+Adj+CL+N] (i.e. with the mass syntactic cue), both of the two groups preferred to interpret them as groups of objects. In other words, when the classifier was *ba*, the adjectives preceding the classifiers forced participants to interpret *ba* as a massifier. As a consequence, the nominal phrases with the Adj-CL word order were interpreted as expressing quantitative and plural sets meanings. This is consistent with the claim from Cheng & Sybesma (1998, 1999) that being modified by adjectives is a syntactic cue for massifiers, and a massifier in a nominal phrase always forces a mass/plural reading. In addition, this is also consistent with the results from Li et al (2008) that nominal phrases with different structures have different interpretations, and the mass-preferred structure would lead to a mass/plural interpretation of the nominal phrase.

When the noun was a mass noun, in Condition 1, there was no preference for either the mass-high-Adj picture or the mass-low-Adj picture. It is reasonable to observe this finding since in Condition 1 where there was no adjective in the nominal structure, the size of the objects in the picture would not affect participants' choice. As long as the picture contains the objects the nouns denoted, they would be chosen. In Condition 2, participants preferred the mass-low-Adj picture as the target picture, while in Condition 3, they preferred the mass-high-Adj picture. Recall that both of these two pictures contain the entities denoted by the same mass nouns. The difference between these two pictures is in the mass-high-Adj picture, the size of the groups of the entities is consistent with the adjective, while in the mass-low-Adj picture, the size of each individual entity is consistent with the adjective, but the size of the group is not. When the nominal phrases have the structure [Num+CL+Adj+N] in Condition 2, in which the adjectives precede and modify the nouns, participants preferred the mass-low-Adj picture in which the size of each individual entity is consistent with the adjectives. When the nominal phrases have the structure [Num+Adj+CL+N] in Condition 3 in which the adjectives precede

and modify the classifiers, participants chose the mass-high-Adj picture as the target picture since the size of the group in this picture is consistent with the adjectives.

When the classifier was *gen/kuai*, different from the situations with the classifier *ba*, no matter whether the noun was a count noun or a mass noun, different nominal structures did not affect either L1ers' or L2ers' interpretations. When the noun was a count noun, in all three conditions, participants chose the count-count picture as the target picture and ignored the count-mass picture even though both of these two pictures contain the objects denoted by the same count nouns. By choosing the count-count picture, participants interpreted the nominal phrase as expressing countable objects with discrete units, while by choosing the count-mass picture, participants interpreted the nominal phrases as expressing quantitative meanings, in other words, objects divided in certain shaped units and counted by the divided units instead of their natural units. The results indicate that no matter what structure the nominal phrases have (CL-Adj vs. Adj-CL), participants always interpreted them as expressing countable objects with their natural discrete units. This is different from my predictions. Based on Cheng & Sybesma (1998, 1999) I predict that mass-biased structures would always force a mass/plural interpretation on the nominal phrases. And also, according to the fixation distributions in Condition 3 with the classifier *gen/kuai*, on hearing the classifiers directly following the adjectives, both L1ers and L2ers landed more fixations on the mass/plural-expressing pictures (the count-mass picture, the mass-high-Adj picture, and the mass-low-Adj picture), which indicates that during real time predictive processing, participants took advantage of the Adj-CL word order to build a mass-biased structure, and expected a mass/plural-expressing interpretation on the nominal phrases. Why did the onset of count nouns in this situation change participants' interpretations to a count-expressing meaning? A possible explanation to this surprising finding is the conflict between the semantic properties of the count nouns and the massifiers. It should be noted that being massifiers, *ba* has a collective semantic meaning while *gen* and *kuai* have the dividing meanings (Cheng & Sybesma, 2010). Recall that, the count nouns and mass nouns used in the current research had been normed by using the definition from Krifka (1989): 'dividability'. Count nouns were rated as objects that cannot be divided or after being divided each piece of it would lose the original properties. Mass nouns were rated as objects that can be divided. When count nouns occurred in Condition 3 with the classifier *gen/kuai* which have the 'divider' meaning, there would be a conflict between the

semantic properties of the massifiers and the nouns, since the massifiers create a certain measurement by dividing the entities denoted by the nouns, while the entities denoted by the nouns cannot be divided. The observation that participants preferred the count-count picture rather than the count-mass picture to be the target in Condition 3 when the classifier was *gen/kuai* indicates that when there is a conflict between the semantic properties of the classifier and the noun, the semantic properties of the nouns would override the classifiers'. As a consequence, in Condition 3, even though the Adj-CL word order would force a massifier reading for *gen/kuai*, participants still chose the count-count picture as the target picture in which the objects denoted by the count nouns are presented in their natural discrete units.

When the noun was a mass noun, in Condition 1, there was no preference for either the mass-high-Adj picture or the mass-low-Adj picture. In both Condition 2 and Condition 3, participants preferred the mass-high-Adj picture as the target picture rather than the mass-low-Adj picture. Recall that both the mass-high-Adj picture and the mass-low-Adj picture contain three portions of entities denoted by the same mass nouns. The only difference between these two pictures is in the mass-high-Adj picture the size of the portion is consistent with the adjective, while in the mass-low-Adj picture, the size of the portion is not. When the audio nominal phrases have the structure [Num+CL+N] in Condition 1, both of the mass-noun denoting pictures can be described by these nominal phrases. In Condition 2 and Condition 3, participants preferred the mass-high-Adj picture in which the size of the portions is consistent with the adjectives, regardless of whether the nominal phrases have the structure [Num+CL+Adj+N] or [Num+Adj+CL+N]. This is because the entities denoted by mass nouns do not have specific discrete units in nature. They can only be counted by different measure units. Hence no matter whether the adjectives are preceding and modifying the classifiers, or preceding and modifying the nouns, it should be the size of the divided units of the entities that is consistent with them.

In general, the finding that L2ers exhibited similar behaviour patterns with L1ers indicates that L2ers interpreted the nominal phrases with different structures the same way L1ers did. Both L1ers and L2ers were sensitive to the Adj-CL word order and its mass-preferred meanings. Also, both of the two groups were familiar with the differences of the massifier meanings of the classifiers (i.e. being massifiers, *ba* has a 'collective' meaning while *gen* and *kuai* have 'divider' meanings). When encountering a conflict

between the semantic properties of the classifier and the noun, both L1ers and L2ers chose the noun over the classifier.

### 6.3. Fixation distributions

The Visual World Paradigm offers us real-time data which provides time-sensitive information about how the onset of each item in nominal phrase affects native Mandarin speakers' and late Dutch-Mandarin learners' predictive processing.

Based on the results of the Cloze Probability Test and the separate analysis of participants' fixation distributions in the analysis windows of classifiers and nouns, a comparison between my predictions and native Mandarin speakers' fixation distributions is summarized in Table 21. Further, a comparison between native Mandarin speakers' and high proficiency Dutch-Mandarin learners' fixation distributions is summarized in Table 22.

*Table 21. A comparison between predictions vs. L1ers' results ('pre' represents 'prediction', 'res' represents 'results', '=' represents no obvious preference for any picture, CC stands for the count-count picture, CM stands for the count-mass picture, MHA stands for the mass-high-Adj picture, and MLA stands for the mass-low-Adj picture. The check mark  $\checkmark$  means L1ers' results are consistent with the predictions. The differences between predictions and L1ers' results, and between L1ers' and L2ers' results are highlighted with shadows)*

When the classifier was <i>ba</i>								
C1	CI		Count-N		Mass-N			
	<i>pre</i>	<i>res</i>	<i>pre</i>	<i>res</i>	<i>pre</i>	<i>res</i>	<i>pre</i>	<i>res</i>
	CC=CM> MHA=MLA	$\checkmark$	CC>all	$\checkmark$	CC=CM< MHA=MLA	$\checkmark$		
C2	CI		Adj		Count-N		Mass-N	
	<i>pre</i>	<i>res</i>	<i>pre</i>	<i>res</i>	<i>pre</i>	<i>res</i>	<i>pre</i>	<i>res</i>
	CC=CM> MHA=MLA	$\checkmark$	CC=CM> MHA=MLA	Na <sup>14</sup> .	CC>all	$\checkmark$	MLA>all	$\checkmark$
C3	Adj		CI		Count-N		Mass-N	
	<i>pre</i>	<i>res</i>	<i>pre</i>	<i>res</i>	<i>pre</i>	<i>res</i>	<i>pre</i>	<i>res</i>
	CC<all	$\checkmark$	CC<all	$\checkmark$	CM>all	$\checkmark$	MHA>all	$\checkmark$
When the classifier was <i>gen/kuai</i>								

<sup>14</sup> Since I did not analysis the fixation distributions in the window of adjective by separating different classifiers, there is no clear data to show L1ers' fixation preference in this specific case.

	CI		Count-N				Mass-N	
	<i>pre</i>	<i>res</i>	<i>pre</i>	<i>res</i>	<i>pre</i>	<i>res</i>	<i>pre</i>	<i>res</i>
C1	Na <sup>15</sup> .	=	CC=CM> MHA=MLA	CC>all	CC=CM< MHA=MLA			√
	CI		Adj		Count-N		Mass-N	
	<i>pre</i>	<i>res</i>	<i>pre</i>	<i>res</i>	<i>pre</i>	<i>res</i>	<i>pre</i>	<i>res</i>
C2	Na <sup>16</sup> .	=	=	Na <sup>17</sup> .	CC=CM> MHA=MLA	CC>all	MHA>all	√
	Adj		CI		Count-N		Mass-N	
	<i>pre</i>	<i>res</i>	<i>pre</i>	<i>res</i>	<i>pre</i>	<i>res</i>	<i>pre</i>	<i>res</i>
C3	CC<all	√	CC<all	√	CM>all	CC>all	MHA>all	√

Table 22. A comparison between L1ers' vs. L2ers' results ('L1ers' represents 'native Mandarin speakers' results', 'L2ers' represents 'Dutch-Mandarin learners' results', '=' represents no obvious preference for any picture, CC stands for the count-count picture, CM stands for the count-mass picture, MHA stands for the mass-high-Adj picture, and MLA stands for the mass-low-Adj picture. The check mark √ means L2ers' results are consistent with L1ers. The differences between predictions and L1ers' results, and between L1ers' and L2ers' results are highlighted with shadows.)

#### When the classifier was *ba*

	CI		Count-N				Mass-N	
	<i>L1ers</i>	<i>L2ers</i>	<i>L1ers</i>	<i>L2ers</i>	<i>L1ers</i>	<i>L2ers</i>	<i>L1ers</i>	<i>L2ers</i>
C1	CC=CM> MHA=MLA	=	CC>all	√	CC=CM< MHA=MLA			√
	CI		Adj		Count-N		Mass-N	
	<i>L1ers</i>	<i>L2ers</i>	<i>L1ers</i>	<i>L2ers</i>	<i>L1ers</i>	<i>L2ers</i>	<i>L1ers</i>	<i>L2ers</i>
C2	CC=CM> MHA=MLA	=	=	MHA>all	CC>all	√	MLA>all	√

<sup>15</sup> Based on the results of the Cloze Probability Test, it is unclear whether the classifiers *gen* and *kuai* are mass-biased classifiers or equally-biased classifiers. Thus it is not possible to make a prediction that which picture participants would prefer on hearing these two classifiers.

<sup>16</sup> Same to the footnote 15.

<sup>17</sup> Same to the footnote 14.

	<b>Adj</b>		<b>Cl</b>		<b>Count-N</b>		<b>Mass-N</b>	
<b>C3</b>	<i>L1ers</i>	<i>L2ers</i>	<i>L1ers</i>	<i>L2ers</i>	<i>L1ers</i>	<i>L2ers</i>	<i>L1ers</i>	<i>L2ers</i>
	<b>CC&lt;all</b>	<b>MHA&gt;all</b>	CC<all	√	CM>all	√	MHA>all	√
<b>When the classifier was <i>gen/kuai</i></b>								
	<b>Cl</b>		<b>Count-N</b>		<b>Mass-N</b>			
<b>C1</b>	<i>L1ers</i>	<i>L2ers</i>	<i>L1ers</i>	<i>L2ers</i>	<i>L1ers</i>	<i>L2ers</i>		
	=	√	CC>all	√	CC=CM< MHA=MLA		√	
	<b>Cl</b>		<b>Adj</b>		<b>Count-N</b>		<b>Mass-N</b>	
<b>C2</b>	<i>L1ers</i>	<i>L2ers</i>	<i>L1ers</i>	<i>L2ers</i>	<i>L1ers</i>	<i>L2ers</i>	<i>L1ers</i>	<i>L2ers</i>
	=	√	=	<b>MHA&gt;all</b>	CC>all	√	MHA>all	√
	<b>Adj</b>		<b>Cl</b>		<b>Count-N</b>		<b>Mass-N</b>	
<b>C3</b>	<i>L1ers</i>	<i>L2ers</i>	<i>L1ers</i>	<i>L2ers</i>	<i>L1ers</i>	<i>L2ers</i>	<i>L1ers</i>	<i>L2ers</i>
	<b>CC&lt;all</b>	<b>MHA&gt;all</b>	CC<all	√	CC>all	√	MHA>all	√

In Table 21, when the classifier was *gen/kuai*, the differences between the predictions and L1ers' results are on hearing count nouns, in Condition 1 and Condition 2, native Mandarin speakers chose the count-count picture instead of choosing between the two count-noun denoting pictures. And in Condition 3, native Mandarin speakers chose the count-count picture instead of the count-mass picture. As will be discussed later, these differences between the predictions and L1ers' results reflect count nouns' 'undividable' property overriding the 'dividing' properties of the classifiers and the structures.

In Table 22, when the classifier was *ba*, native Mandarin speakers and Dutch-Mandarin learners behaved differently to the classifier and the adjectives. As will be discussed later, these differences reflect L2ers' insensitivity to the count-bias property of the classifier *ba*, and their different processing patterns of adjectives and the Adj-Cl word order. When the classifier was *gen/kuai*, L2ers' special processing patterns of adjectives and the Adj-Cl word order are also observed.



In this part, participants' general fixation distribution patterns in the three conditions are discussed first, followed by the discussions of the fixation distributions in the analysis window of the classifier and the analysis window of the noun.

### *6.3.1. General fixation distribution patterns*

By comparing participants' fixation distributions in Condition 1 with Condition 2 we can see that, adding an adjective to modify the noun did not affect L1ers' on-line processing and anticipation building. In both conditions they started to fix on the possible target pictures until the onset of the nouns. As for L2ers, the onset of the adjectives following the classifiers drove their attention to the mass-high-Adj picture in which the size of the portions is consistent with the adjectives. By comparing participants' fixation distributions in Condition 2 with Condition 3 we can see that, the word order of adjectives and classifiers had an effect on participants' fixation distributions. When the adjectives precede the classifiers (in Condition 3 with the structure [Num+Adj+CL+N]), L1ers started to drive their attention to the mass/plural-expressing pictures on hearing the adjectives. While in Condition 2 where the adjectives follow the classifiers ([Num+CL+Adj+N]), the onset of the adjectives had no effect on L1ers' fixation distributions. These observations are consistent with the claim from Cheng & Sybesma (1998) that being modified by adjectives is a syntactic cue for massifier readings. As for L2ers, similar to their fixation distributions in Condition 2, in Condition 3 on hearing the adjectives following the numeral, they shifted more attention to the mass-high-Adj picture. However, on hearing the classifiers following the adjectives, they started to fix more on the mass/plural-expressing pictures (the count-mass picture, the mass-high-Adj picture, and the mass-low-Adj picture) just like L1ers.

The general fixation distribution patterns indicate that L2ers exhibited native-like fixation distribution patterns in Condition 1, but non-native-like patterns in Condition 2 and Condition 3. To be specific, the adjectives in nominal phrases affected L1ers' and L2ers' predictive processing differently. For L1ers, the onset of the adjectives occurring after the classifiers (in Condition 2) did not affect their real time processing. Their fixation distributions in the time window of the adjective were along the same trend as in the time window of the classifier. However, the onset of the same adjectives occurring directly after the numeral (in Condition 3) drove their attention to the mass/plural-expressing pictures (the count-mass picture, the mass-high-Adj picture, and the mass-low-Adj picture). As for L2ers, the onset of the adjectives drove their attention to the

mass-high-Adj picture regardless of the word order of the adjectives and the classifiers. In both Condition 2 and Condition 3, on hearing the adjectives, L2ers started to fix on the mass-high-Adj picture in which the size of the groups of the objects is consistent with the adjectives. This difference between L1ers and L2ers indicates that L1ers can use the adjectives directly following the numeral to build a mass-biased structure immediately, and then expect the nominal phrase to have a mass/plural-expressing meaning before finishing reading the nominal phrases. This explains L1ers' significantly more fixations in the mass/plural-expressing pictures (the count-mass picture and the two mass-noun denoting pictures) than in the count-expressing picture (the count-count picture) on hearing the adjectives directly following the numeral which is far before the onset of the nouns. On the other hand, L2ers exhibited their sensitivity to the lexical-semantic meanings of the adjectives, and always focused on the pictures in which the size of the groups of the objects is consistent with the adjectives, but ignored the syntactic function of the adjectives when they occurred directly following the numeral (i.e. the mass-preference of the Adj-CL word order). This observation indicates that compared to L1ers who were more sensitive to the syntactic information of the adjectives (the mass-preference of the Adj-CL word order), L2ers were more sensitive to the lexical-semantic information of them (the Adj-consistent size).

Furthermore, L2ers differed from L1ers in how quick they use the mass syntactic cue (Adj-CL word order) to predict mass/plural-expressing interpretations. In Condition 3 where the nominal phrases have the structure [Num+Adj+CL+N], the onset of the adjectives directly following the numeral led L1ers' attentions to the mass/plural-expressing pictures, while L2ers only drove their fixations to these pictures until the onset of the classifiers following the adjectives. Even though both of the two groups fixed more on the mass/plural-expressing pictures than the count-expressing picture in the time window of the classifier, L1ers did so since the onset of the adjectives which is earlier than L2ers. This finding indicates that even though both L1ers and L2ers were sensitive to the Adj-CL word order and can take advantage of this syntactic cue to make mass/plural-expression predictions, L2ers were slower than L1ers in making use of this cue. Compared to L1ers, L2ers need longer time and more information to build anticipations during real time processing.

### 6.3.2. *The analysis window of the classifier*

Similar to the behaviour results, the ‘classifier division’ effect has also been found in on-line fixation distributions in the analysis window of the classifier on L1ers. In the analysis window of the classifier, the classifiers *gen* and *kuai* had similar fixation distribution patterns with each other, which were different from the classifier *ba*’s: in Condition 1 and Condition 2 (where the nominal phrases have neutral structures), the onset of the classifier *ba* evoked L1ers’ attention to the count-noun denoting pictures, while the onset of the classifier *gen/kuai* did not affect participants’ fixation directions. This observation is consistent with the results of the Cloze Probability Test that count nouns were preferred to occur following the classifier *ba*, while the classifiers *gen/kuai* exhibited no preference for either count nouns or mass nouns. Also, it is reasonable to observe a count-noun-preference on the processing of the classifier *ba* if we consider it as a count-biased classifier.

However, if we treat *ba* as a count-biased classifier which has a dominant count-classifier meaning, one would expect significantly more fixations in the count-count picture than in the count-mass picture on hearing the classifier *ba* in Condition 1 and Condition 2, since the dominant count-classifier meaning of *ba* can only be reflected through the single individual units presented in the count-count picture but not the count-mass picture in which objects are organized in groups. This is different from the observation that in the neutral structures (Condition 1 and Condition 2), the onset of *ba* triggered fixations in both the count-count picture and the count-mass picture. The difference of fixation proportions in the analysis window of the classifier between these two pictures was not significant. From my perspective, if we consider the procedure of the Visual World Paradigm experiment, this unexpected finding cannot be treated as the evidence against the result from the Cloze Probability Test that *ba* has a dominant count-classifier meaning. In the Visual World Paradigm experiment, four pictures were presented on the screen at the same time of the onset of the audio material *Qing cong ping mu shang de si fu tupian zhong xuan ze* + nominal phrase ‘From the four pictures on the screen, please choose + nominal phrase’, which means before participants reach the critical nominal phrases, they have around three to four seconds to familiarize with the four pictures. According to previous studies (Seidenberg, Tanenhaus, Leiman & Bienkowski, 1982; Swinney, 1979; Tanenhaus, Leiman & Seidenberg, 1979), all possible meanings of an ambiguous word are accessed initially when such a word is encountered, even though

some of the possible meanings are less-possible. The present of the four pictures (especially the count-mass picture and the two mass-denoting pictures) before the onset of the nominal phrase may activate both the less-dominant massifier meaning and the dominant count-classifier meaning of *ba*. Thus on hearing *ba*, both count-classifier meaning and massifier meaning were activated. Since *ba* has strong connections with associated count nouns, participants at this time point started to fix more on the two pictures containing count objects (the count-count picture and the count-mass picture) than the two mass-denoting pictures, and ignored the difference between the count-count picture and the count-mass picture.

Regarding L2ers, different from the behaviour results, they exhibited some non-native-like patterns in fixation distributions in the analysis window of the classifier. The non-native-like behaviours on L2ers were spotted on their reactions to the count-biased classifier *ba*. As mentioned before, in the analysis window of the classifier, on hearing *ba*, L1ers started to fix more on the count-noun denoting pictures in both Condition 1 and Condition 2, while L2ers did not exhibit any specific preference. Recall that the classifier *ba* is a count-biased classifier based on the results in the Cloze Probability Test. L1ers took advantage of the count-bias property of *ba*, and expected count nouns to be the upcoming nouns on the onset of *ba* which is before the onset of the nouns. This observation indicates that native Mandarin speakers can on-line use the count-bias property of the classifier *ba* to form predictions for the upcoming nouns. L2ers, however, did not exhibit this sensitivity to the count-bias property of *ba*. On hearing *ba* in neutral structures (Condition 1 and Condition 2), L2ers landed their fixations randomly among the four pictures without any preference for either count-noun denoting pictures or mass-noun denoting pictures.

Recall that in Chapter 2, it has been discussed that the semantic properties and the classifier-noun restrictions are treated as lexical information of classifiers. The result here that L2ers were not sensitive to the count-bias property of the classifier *ba* during on-line processing indicates that high proficiency Dutch-Mandarin learners has not fully acquired the lexical information of the classifiers yet.

### 6.3.3. *The analysis window of the noun*

Based on the results in the analysis window of the noun we can see that, among all three conditions, on hearing the count nouns, participants started to look at the count-noun

denoting pictures, while on hearing the mass nouns, participants drove their attention to the mass-noun denoting pictures. This observation indicates that the audio input had an influence on participants' fixation distributions. Also, the fixation distributions in the four pictures are consistent with participants' behaviour data: in the end of each trial, the picture with the highest fixation proportions was chosen to be the target picture by both groups of participants.

The division between the classifier *ba* and the classifiers *gen/kuai* has also been observed on both L1ers and L2ers in the analysis window of the noun. When the classifier was *ba*, the onset of count nouns in Condition 2 triggered more fixations in the count-count picture, while in Condition 3 triggered more fixations in the count-mass picture. The onset of mass nouns in Condition 2 drove more attention to the mass-low-Adj picture while in Condition 3 drove more attention to the mass-high-Adj picture. When the classifier was *gen/kuai*, the onset of the count nouns led more fixations to the count-count pictures in both Condition 2 and Condition 3, and the onset of the mass nouns led more fixations to the mass-high-Adj picture in both Condition 2 and Condition 3. This is consistent with participants' behaviour results.

Comparing participants' fixation distributions and their behaviour data, an interesting finding is revealed: L2ers exhibited native-like behaviours in their final picture decisions (the behavioural results) but not in their fixation distributions. To be specific, the 'classifier division' effect for L2ers was only spotted in the behavioural data, but not in the fixation distributions. This is a reasonable observation if we treat the ambiguity types of classifiers, the restrictions of classifier-noun pairings, and the massifier meanings as different kinds of knowledge. L2ers had not acquired the ambiguity types of classifiers and the restrictions of classifier-noun pairings, thus it is difficult for them to exhibit the sensitivity to the count-bias property of the classifier *ba*. However, they had acquired the massifier meanings (*ba* has a collective meaning while *gen* and *kuai* have the divider meanings), which help them exhibit native-like behaviours in the behaviour results.

On top of that, there comes another question: why were L2ers sensitive to the mass-biased structures (i.e. the Adj-Cl word order) and the massifier meanings, but not sensitive to the ambiguity types of classifiers? All the knowledge about the classifier system and the mass/count syntactic cues in Mandarin (e.g. the ambiguous types of classifiers, classifier-noun restrictions, dual-role classifiers' ambiguous meanings) can

only be implicitly learned by L2ers through immersive language input. Further research needs to be done to explore the reason of this interesting finding.

Taken together, in the Visual World Paradigm experiment I found that, similar to native Mandarin speakers, high proficiency late Dutch-Mandarin speakers can use the Adj-CL word order to build a mass/plural-biased anticipation during real time processing, even though they were slower than natives Mandarin speakers in assigning the mass/plural meaning to the Adj-CL structure. In the current study, high proficiency late Dutch-Mandarin learners can acquire this mass/count syntax cue which is a unique-to-L2 construction and can only be acquired through implicit learning. The results in the Visual World Paradigm experiment support the Feature Reassembly Hypothesis.

The aim of the current study is to explore how different nominal phrase structures affect native Mandarin speakers' and late Dutch-Mandarin learners' predictive processing, thus time-sensitive and space-sensitive information is needed. As discussed in Chapter 3, apart from the Visual World Paradigm, the Reading for Comprehension task is ideal for the current research since it can offer multiple levels of indexes which can be used to analyse people's early as well as late stages of on-line processing (Frenck-Mestre, 2005; Rayner, 1998). The Reading for Comprehension task has been successfully employed in plenty of studies on real time sentence processing (e.g. Blumenfeld & Marian, 2011; Duyck, Van Assche, Drieghe, & Hartsuiker, 2007; Felser, Sato, & Bertenshaw, 2009; Flecken, 2011; Yao & Chen, 2016). It can thus be treated as a well-motivated complement to the Visual World Paradigm to build a more complete picture of how native Mandarin speakers and late Dutch-Mandarin learners comprehend nominal phrases in real time.

Also, in order to look into how tasks with different workloads affect late L2ers' on-line processing, experimental paradigms with different workloads (high vs. low) need to be used on the same group of participants (or at least participants with similar working memory capacities). As discussed in Chapter 3, compared to the Visual World Paradigm, the Reading for Comprehension task needs a higher workload. By comparing high proficiency late Dutch-Mandarin learners' on-line processing in the Visual World Paradigm experiment and the Reading for Comprehension experiment, one can tell whether L2ers' on-line performance is limited by their cognitive resources or deficient

representations (the Fundamental Similarity Approach vs. the Fundamental Difference Approach). The detailed information of the Reading for Comprehension experiment is presented in Chapter 5.

## V. Reading for Comprehension Experiment

The main goal of the current research is to look into whether high proficiency late L2 learners could have native-like behaviours in real time predictive using of unique-to-L2 constructions. As reported in Chapter 4, a Visual World Paradigm experiment was conducted to explore high proficiency late Dutch-Mandarin learners' real time anticipation building by using the mass/count structures (i.e. CL-Adj vs. Adj-CL). The results show that even though slower than native Mandarin speakers, late Dutch-Mandarin learners can on-line use the mass syntactic cue to build mass/plural-expressing anticipations. Some surprising but interesting findings have also been spotted in the Visual World Paradigm experiment. Apart from the mass/count structures, different classifiers' massifier meanings, ambiguity types and the restrictions of classifier-noun pairings exhibited some influences on native Mandarin speakers' predictive processing. Late Dutch-Mandarin learners exhibited a native-like master over the mass/count syntax cue and the massifiers meanings, but failed to show sensitivities to the ambiguity types of classifiers and the restrictions of classifier-noun pairings.

In order to further explore high proficiency late Dutch-Mandarin learners' and native Mandarin speakers' on-line processing of nominal phrases with different structures and different classifiers, a Reading for Comprehension experiment was conducted which requires a higher workload compared to the Visual World Paradigm experiment. By comparing L2ers' on-line processing behaviours between the Reading for Comprehension experiment and the Visual World Paradigm experiment, one can tell whether their on-line performance is restricted by their limited cognitive resources or deficient representations. In addition, as discussed in Chapter 3, based on the results in Li et al. (2008), it is not clear which part of the mass-biased structure [Num+Adj+CL+*de*+N] lead to the mass/plural interpretations. The effect of the CL-Adj/Adj-CL word order has been investigated in the Visual World Paradigm experiment. In this chapter, both the CL-Adj/Adj-CL word order and the insertion of the modification marker *de* were manipulated. Nominal phrases in this current experiment have three types of structures: [Num+CL+Adj+N], [Num+Adj+CL+N], and



[Num+Adj+CL+*de*+N]<sup>18</sup>. Based on the results of the Cloze Probability Test reported in Chapter 4, the classifier *ba* is a count-biased classifier. In this case, putting a count-biased classifier (e.g. *ba*) in mass-preferred structures (e.g. [Num+Adj+CL+N], [Num+Adj+CL+*de*+N]) may cause some processing difficulties since there is a violation between the dominant meaning of the classifier and the preferred meaning of the structure. This possibility is tested in this chapter.

This chapter is structured as follows. The experimental materials are presented in Section 1. Participants information, and the experimental procedure are introduced in Section 2 and Section 3 respectively. Predictions are generated in Section 4, followed by the results in Section 5. General discussions are presented in Section 6.

## 1. Material

Four dual-role classifiers were used in the present study: *gen* ‘rod’, *kuai* ‘chunk’, *pian* ‘slice’, and *ba* ‘handful’. As mentioned in Chapter 1, all of them have two possible interpretations: the count-classifier reading and the massifier reading. In order to examine the influence of the two syntactic cues (i.e. the Adj-Cl word order, and the insertion of the modification marker *de*) on participants’ on-line processing, minimal comparisons were made in the nominal phrase structures. By manipulating both of the two syntactic cues, there are four possible types of nominal phrases, as illustrated in (39).

- (39) a. *yi pian xiao CD*  
           One CL-slice small CD
- b. *yi pian de xiao CD*  
           one CL-slice DE small CD
- c. *yi xiao pian CD*  
           one small CL-slice CD

<sup>18</sup> Based on the results in the Visual World Paradigm experiment that adding an adjective to modify the nouns does not affect participants’ on-line processing of nominal phrases. Thus the baseline structure [Num+CL+N] is not included in the Reading for Comprehension experiment.

d. yi    *xiao*    *pian*    *de*    CD

one   *small*   *CL-slice*   *DE*   CD

The four types of nominal phrases in (39) consist of the same numeral, the same classifier, the same adjective, and the same noun, but with different structures. In (39a) and (39c), there is no *de* inserted after the classifier *pian* ‘slice’, while in (39b) and (39d), the modification marker *de* is inserted after the classifier *pian* ‘slice’. In (39a) and (39b), the adjective *xiao* ‘small’ follows the classifier *pian* ‘slice’, while in (39c) and (39d), the adjective *xiao* ‘small’ precedes the classifier *pian* ‘slice’. Through the contrasts between (39a) and (39b), and the contrast between (39c) and (39d), the influence of the insertion of *de* can be revealed. Through the contrasts between (39a) and (39c), and (39b) and (39d), the effect of the word order of the classifiers and the adjectives can be revealed.

Both count nouns and mass nouns were used in the current experiment, which were chosen based on their typicality of mass/count from the Noun Rating Test reported in Chapter 4. Based on the results of Noun Rating Test, 14 mass nouns and 28 count nouns were chosen for the experimental materials (Presented in Appendix A.1). Nine count nouns and two mass nouns were matched with the classifier *ba* ‘handful’, eight count nouns and four mass nouns were matched with the classifier *gen* ‘rod’, six count nouns and two mass nouns were matched with the classifier *pian* ‘piece’, and five count nouns and six mass nouns were matched with the classifier *kuai* ‘chunk’. The criteria for choosing these nouns were: 1). they have to be typical count/mass nouns; 2). they can be modified by the adjectives *da* ‘big’ or *xiao* ‘small’.

Each nominal phrase is embedded in a simple sentence, with the structure of ‘there is an A in B, B is ...’, in which A refers to the critical nominal phrases and B is a random but reasonable location phrase. Participants’ reading times on each word of the critical nominal phrases were recorded using the eye-tracking software, which can reflect participants’ on-line processing (Frenck-Mestre, 2005; Rayner, 1998, etc.). Even though there could be four types of nominal phrases by manipulating the two syntactic cues, not all of them are grammatical and acceptable at the same extent to native Mandarin speakers. In order to make sure that all the nominal phrases and sentences in the current experiment are well-formed and grammatical to native Mandarin speakers, a Grammaticality Test was conducted before the Reading for Comprehension experiment.

### 1.1. Grammaticality Test: The classifier-noun pairs and the sentences

In order to make sure that all the classifier-noun pairs in the current experiment are matched, and all the material sentences are grammatical and acceptable to native Mandarin speakers, in the Grammaticality Test, 20 native Mandarin speakers (who had not participated in either the Norming Tests or the two critical eye-tracking experiments) were asked to rate all the sentences on a 5-point scale based on how acceptable the sentences are: 1 stands for unacceptable, while 5 stands for acceptable. All sentences were pseudo-randomly divided into four lists, each list only contained one of the four types of each nominal phrase. Each participant read one of the four lists.

The results from the Grammaticality Test showed that the mean rating scores of sentences with the nominal phrase structures [Num+CL+Adj+N], [Num+Adj+CL+N], and [Num+Adj+CL+*de*+N] were around 4.5, which means these sentences are grammatical and acceptable to native Mandarin speakers. The average score for the sentences with the structure [Num+CL+*de*+Adj+N] was around 3.0, indicating that native Mandarin speakers did not like this type of nominal phrases. Some of the native speakers reported that they could understand roughly what the sentences with the nominal phrase [Num+CL+*de*+Adj+N] are trying to express, but the sentences are odd to them. Some others reported that they had difficulties to get the meanings of the sentences in this structure. Given that the purpose of the current study is to look into participants' on-line processing of matched classifier-noun pairs and acceptable sentences, the sentences with the nominal phrase structure [Num+CL+*de*+Adj+N] were excluded from the materials.

Based on the results from Grammaticality Test, there were three types of nominal phrase structures for each classifier-noun pair (T is short for Type): T1 with the structure [Num+CL+Adj+N], T2 with the structure [Num+Adj+CL+N], and T3 with the structure [Num+Adj+CL+*de*+N]. Recall that there were 14 mass nouns and 28 count nouns, thus in total there were 126 sentences in the materials  $((14+28) * 3)$ . All the sentences were pseudo-randomly divided into three lists to make sure that for each classifier-noun pair, the three types of nominal phrases occurred in three different lists. Each list also contained 50 fillers. Two adjectives were used in the materials: *da* 'big' and *xiao* 'small'. In each list, half of the sentences were with the adjective *da* 'big', and the other half were

with the adjective *xiao* ‘small’. Examples of the material sentences can be seen in Table 23, and all material sentences can be seen in Appendix A.2.

Table 23. Examples of the material sentences

Interest Regions (each colour represents one Interest Region)		
<i>Hezi li zhuang zhe yi</i> Box inside put zhe one	<i>ba xiao/da shanzi,</i> CL small/big fan,	<i>na hezi shi hong se de.</i> that box is red colour DE.
	<i>xiao/da ba shanzi,</i> small/big CL fan,	
	<i>xiao/da ba de shanzi,</i> small/big CL de fan,	
‘There is a small (big) fan/ There is a small (big) handful of fans in the box, the box is red. ’		

From Table 23 we can see that, the critical nominal phrase is in the middle of the whole sentence, but at the end of the first internal-clause with a comma following it. Some previous eye-tracking studies found that punctuations (e.g. comma, period) have some influences on readers’ reading times on the chunk before the punctuation (i.e. the wrap-up effect, see Just & Carpenter, 1980; Rayner, Sereno, Morris, Schmauder, & Clifton, 1989). Recent studies found that even though both sentence complexity and punctuation have effects on sentence parsing, there is no interaction between them in any eye-movement measure (Warren, White, & Reichle, 2009). The sentence complexity (nominal phrase structures) is the key research question in the current study considering that the aim of the current experiment is to explore the nominal phrase structures’ effects on L1ers’ and L2ers’ on-line processing. Since the critical nominal phrases with different structures are all in the same position of the clause, and the punctuation and the structure complexity would not interact, the comma would cause a same wrap-up effect on all the phrases and the nouns. In this case, the possible wrap-up effects on all the phrases and the nouns would be balanced through the deductions of processing times between any two conditions. Taken together, the comma following the critical nominal phrases in the material is not a concern in the current experiment.

To sum up, there were two factors in the materials: the insertion of the modification marker *de* and the CL-Adj word order. By manipulating both of these two factors to have two levels (nominal phrases with or without *de* being inserted after the classifier, and

nominal phrases with CL-Adj or Adj-CL order), there were four possible types of nominal phrases. Based on the results of the Grammaticality Test, only three out of the four possible types of nominal phrases were chosen in the current study. The manipulation of these two factors in current study can be seen in Table 24.

*Table 24. The manipulation of the two factors*

	With <i>de</i>	Without <i>de</i>
CL + Adj	-	T1
Adj + CL	T3	T2

## 2. Participants

Two groups of participants took part in the Reading for Comprehension experiment: native Mandarin speakers (L1ers) and high proficiency late Dutch-Mandarin learners (L2ers).

The native group consisted of 50 students (33 females and 17 males) from Beijing Normal University, all of them are native speakers of Mandarin with normal or corrected-to-normal vision (aged between 18 to 28 years old). Each native participant was given ¥ 50 for their participation. The data of the native group was collected in Beijing Normal University, Beijing, China.

The L2 learners were the same high proficiency late Dutch-Mandarin learners who also participated in the Visual World Paradigm experiment. The data of the L2er group was collected in Leiden University, Leiden, Netherland. As reported in Chapter 4, in order to reduce the possible priming/learning effect on L2ers, the order of the two eye-tracking experiments and off-line tests was manipulated: the Naming Test → the Reading for comprehension experiment/the Visual World Paradigm experiment → the Blank-Filling Test → the Visual World Paradigm experiment/the Reading for comprehension experiment. There was a 10-minutes break between any two tests.

## 3. Procedure

In a Reading for Comprehension task based on the eye-tracking technique, the first fixation duration could help us gain a clear picture of the very early stage processing of the classifiers and the nouns, and the predictability of a certain classifier or a noun in a specific nominal structure. The total reading time could offer us information about the late stage of on-line processing and the possible difficulties of understanding a certain region. In addition, the Reading for comprehension task is asking participants to read the sentences for comprehension: participants are required to focus on the meaning of the sentences instead of the grammar, which could help us to look into participants' unconscious processing of nominal phrases, excluding all the potential explicit influences.

In the current study, the SR Research Eyelink 1000 was used to measure participants' eye movements. Both left and right eye movements were recorded, but only the right eye data was analysed. Following calibration and validation, gaze-position error was less than 0.5°. All sentences were displayed in one line in the middle of the screen, and the numbers of characters in one sentence varied from 17 to 24. Participants were seated 60cm from a 19-inch monitor.

Participants were tested individually. Before the critical experimental trials, there was a brief instruction and 10 practice trials. After participants read the instruction and finished the practice trials, a standard 9-point grid calibration and validation were completed. During the critical experimental trials, participants were required to read the sentence on the screen silently. When finished reading, participants need to click the mouse so that the original sentence will disappear and be replaced by a meaning-related sentence. Participants were asked to make a judgement about whether this current sentence is consistent with the previous one in meaning, and move the mouse to click the corresponding button to respond. At the beginning of each trial, a drift calibration screen appeared, and participants were asked to try their best to fix on the calibration dot which was presented in the middle of the left side of the screen. A trial only started when participants fixed on the calibration dot stably. The ethical protocol approval was obtained from Beijing Normal University and Queen Mary Ethics of Research Committee.

## 4. Predictions

According to Cheng & Sybesma (1998, 1999), being modified by adjectives and the insertion of *de* are two syntactic cues for massifiers, which would lead to the anticipation of mass or plural meanings. Also according to Li et al. (2008), nominal phrases with the structure [Num+Adj+CL+*de*+N] will lead participants to choose mass/plural-expressing objects. In the current study, among the three types of nominal phrases, T2 ([Num+Adj+CL+N]) and T3 ([Num+Adj+CL+*de*+N]) are mass/plural-preferred structures while T1 ([Num+CL+Adj+N]) is a neutral structure which has no preference for either count or mass meanings. Based on the results in Cloze Probability Test reported in Chapter 4, *ba* is a count-biased classifier which has a dominant count-classifier meaning and a close connection with associated count nouns. On the other hand, the classifiers *gen/kuai* are non-count-biased classifiers, they could be either mass-biased or equally-biased classifiers. The classifier *pian* is also tested in this Cloze Probability Test. The results showed that similar to the classifiers *gen/kuai*, *pian* is a non-count-biased classifier: the overall cloze probabilities of count nouns and mass nouns following *pian* were similar to each other (46.67% vs. 53.33%). The results of Cloze Probability Test including the classifier *pian* is illustrated in Table 25 and Table 26.

Table 25. The cloze probability of count and mass nouns

Classifier	Count nouns				Mass nouns			
	1 <sup>st</sup> (%)	2 <sup>nd</sup> (%)	3 <sup>rd</sup> (%)	Overall	1 <sup>st</sup> (%)	2 <sup>nd</sup>	3 <sup>rd</sup> (%)	Overall
<i>Ba</i>	95	87.5	75	<b>85.83</b>	5	12.5	25	14.16
<i>Gen</i>	55	40	50	48.4	45	60	50	51.6
<i>Kuai</i>	47.5	55	45	49.17	52.5	45	55	50.83
<i>Pian</i>	52.5	40	47.5	46.67	47.5	60	52.5	53.33

Table 26. The interpretation of nominal phrases (%)

	<i>Ba</i>		<i>Gen</i>		<i>Kuai</i>		<i>Pian</i>	
	Count N	Mass N	Count N	Mass N	Count N	Mass N	Count N	Mass
<b>1</b>	99.16	0.79	98.77	90.53	99.49	87.69	98.38	48.11
<b>?</b>	0.84	99.21	1.23	9.47	0.51	12.31	1.62	51.89

The co-occurrence frequency of count nouns and mass nouns following the four classifiers are summarized in Table 27.

Table 27. The co-occurrence frequency of count nouns and mass nouns (per million)

	<b><i>Ba</i></b>			<b><i>Gen</i></b>		
	Count N	Mass N	<i>diff</i>	Count N	Mass N	<i>diff</i>
<b>Frequency</b>	0.689	0.641	-0.048	0.356	0.513	0.157*
	<b><i>Kuai</i></b>			<b><i>Pian</i></b>		
	Count N	Mass N	<i>diff</i>	Count N	Mass N	<i>diff</i>
<b>Frequency</b>	0.432	0.594	0.162*	0.371	0.605	0.234*

Thus, the differences among the four classifiers are summarized in Table 28.

Table 28. The differences among the four classifiers

	Massifier meaning	Ambiguity type	Preferred nouns	Co-occurrence frequency
<b><i>Ba</i></b>	Collective	Count-biased	Count nouns	Count = Mass
<b><i>Gen</i></b>	Divider	Mass/Equal-biased	Count/Mass nouns	Count < Mass
<b><i>Kuai</i></b>	Divider	Mass/Equal-biased	Count/Mass nouns	Count < Mass
<b><i>Pian</i></b>	Divider	Mass/Equal-biased	Count/Mass nouns	Count < Mass

Among the four dual-role classifiers used in the current study, *ba* is a count-biased classifier while *gen/kuai/pian* are non-count-biased classifiers. Considering the mass/count-biased structures and the mass/count biased classifiers, an interaction between these two factors is expected. According to previous studies (Rayner & Duffy, 1986, etc.), prior context has an effect on the real time processing of ambiguous words (cf. Libben, 2006; Pollatsek, Drieghe, Stockall, & De Almeida, 2010). To be specific, when an ambiguous word has imbalanced possible meanings and the prior context favour the less dominant meaning, the reading times should be slow on this word compared to when the prior context favour the dominant meaning (the Subordinate Bias



Effect, Duffy, Morris, & Rayner, 1988; Rayner & Duffy, 1986; Sereno, Brewer, & O'Donnell, 2003; Sereno, O'Donnell, & Rayner, 2006).

In the current study, I predict that different types of ambiguous words would behave differently with different nominal phrase structures. To be specific, putting a count-biased classifier (i.e. *ba*) in a mass-biased structure (e.g. [Num+Adj+CL+N], [Num+Adj+CL+*de*+N]) will cause some processing difficulties since there is a conflict between the preferred meaning from the structure and the dominant meaning of the classifier. In this case, the classifier *ba* is expected to have longer fixation durations in T2 and T3 than in T1, since in T2 and T3 where the classifier is being modified by adjectives, the adjective preceding the classifier can be treated as a context cue for the massifier reading. Leading by this cue, participants have to suppress the dominant count-classifier reading of the count-biased classifier and choose the less-likely massifier reading to be the final reading. The procedure of suppressing the dominant reading and forcing the less-likely reading to participate in a higher level of processing would take extra processing time. When the classifier is a mass-biased classifier, the processing time on it in the mass-biased nominal phrase (i.e. T2 and T3) should have no difference with that in the neutral nominal phrase (T1), since the dominant massifier meaning can fit in all the three nominal types. When the classifier is an equally-biased classifier, the processing time on it in the biased nominal phrases (T2 and T3) should be shorter than that in the neutral nominal phrase (T1), since the equally possible meanings of an equally-biased classifier would cause some competitions during processing in the neutral nominal phrase, but the syntactic cue in the biased nominal phrase would eliminate this processing difficulty and facilitate the processing by picking one of the possible meanings to be the final one (Rayner & Duffy, 1986). Also, Swinney (1979) and Seidenberg et al. (1982) found evidence that, although all the possible meanings of an ambiguous word are accessed initially, readers usually make a decision and select a meaning as the final reading of an ambiguous word within 200ms even in the absence of a disambiguating context. In the current study, the first fixation durations on the classifiers could be the index reflecting participants' on-line processing of different types of ambiguous classifiers. The prediction of the first fixation durations on the classifiers in nominal phrases with different structures are summarized in Table 29.

Table 29. The predictions of processing times on classifiers

		Count-biased classifier	Mass-biased classifier	Equally- biased classifier
T1[Num+CL+Adj+N]	Neutral			
T2[Num+Adj+CL+N]	Mass- biased	$T1 < T2 = T3$	$T1 = T2 = T3$	$T1 > T2 = T3$
T3[Num+Adj+CL+ <i>de</i> +N]	Mass- biased			

Table 30. The predictions of processing times on nouns

	<i>Ba</i>		<i>Gen/Kuai/Pian</i>	
	Count N	Mass N	Count N	Mass N
T1[Num+CL+Adj+N]				
T2[Num+Adj+CL+N]	$T1 < T2 = T3$	$T1 > T2 = T3$	$T1 = T2 = T3$	
T3[Num+Adj+CL+ <i>de</i> +N]			Count > Mass	

Also, different classifiers have different restrictions with their associated nouns. As illustrated in Table 28, the classifier *ba* prefers count nouns than mass nouns, while the classifiers *gen/kuai/pian* are happy to occur with either count nouns or mass nouns. Thus count nouns occurring after *ba* are expected to have shorter processing times than mass nouns (Ehrlich & Rayner, 1981). Moreover, count nouns occurring with *ba* in T1 should have shorter processing times than in T2 and T3, since in T1 the count-classifier dominant meaning of *ba* would facilitate the processing of its associated count nouns. The mass nouns, on the other hand, should have longer processing times with *ba* in T1 than in T2 and T3, since the count-classifier dominant meaning of *ba* in T1 cannot occur with mass nouns. On reaching the mass nouns in T1, participants need to suppress the dominant count-classifier meaning of *ba* and activate the less-likely massifier meaning. When the classifier is *gen/kuai/pian*, the processing times of count nouns and mass nouns should have no difference among the three types of nominal phrase structures, since no matter in which type of nominal phrase, the classifier will always be a massifier, which can occur with either count nouns or mass nouns. Also, based on the co-occurrence frequency of count nouns and mass nouns showed in Table 24, count nouns which have lower frequency compared to mass nouns should have longer fixation times than mass nouns when the classifiers are *gen/kuai/pian*. The predictions

of the processing times on the nouns following different classifiers are summarized in Table 30.

In general, the predictions for native Mandarin speakers' first fixation duration patterns<sup>19</sup> on classifiers and nouns are summarized in Table 31. (FFD is short for First Fixation Durations. The numbers 1,2, and 3 represent the three nominal phrase types T1, T2, and T3 respectively. FFD1 stands for the First Fixation Durations on classifiers in T1, and so on. FFDcount represents the First Fixation Durations on the count nouns, while FFDmass represents the First Fixation Durations on the mass nouns.)

Table 31. Predictions about L1ers' first fixation durations on classifiers and nouns

<b>FFD on classifiers</b>	<b><i>Ba</i></b>	FFD1<FFD2=FFD3	<b><i>Ba</i></b> is a count-biased classifier
	<b><i>Gen/</i></b>	FFD1=FFD2=FFD3	<b><i>Gen/Kuai/Pian</i></b> are mass-biased classifiers
	<b><i>Kuai/</i></b>		
	<b><i>Pian</i></b>	FFD1>FFD2=FFD3	<b><i>Gen/Kuai/Pian</i></b> are equally-biased classifiers
<b>FFD on nouns</b>	<b><i>Ba:</i></b>		<b><i>Ba</i></b> is a count-biased classifier which has
	Count N: FFD1<FFD2=FFD3		strong connections with count nouns; count-
	Mass N: FFD1>FFD2=FFD3		classifiers only occur with count nouns
	<b><i>Gen/</i></b>	FFDcount>FFDmass	Frequency effect
	<b><i>Kuai/</i></b>		
	<b><i>Pian</i></b>	FFD1=FFD2=FFD3	Massifiers can occur with either count or mass nouns

As for high proficiency Dutch-Mandarin learners, as have discussed in Chapter 2 and based on the results in the Visual World Paradigm Experiment reported in Chapter 4, in order to accomplish the tasks in the current research, they need to acquire not only the syntactic knowledge of the count/mass classifiers (i.e. the Adj-Cl word order, and the

<sup>19</sup> I did not raise the prediction for native Mandarin speakers' total reading times because the total reading times not only reflect participants' late stage processing, but also are affected by first fixation duration, regression and some other processing index. It is difficult to have a very precise prediction about participants' total reading times.

insertion of *de*), but also the lexical knowledge of them (i.e. the ambiguity type, the relations with preferred count/mass nouns). Thus, there are four possibilities for Dutch-Mandarin learners' behavioural patterns, which are summarized in Table 32.

Table 32. Predictions about L2ers' behaviours

L2ers have already acquired...	Their behavioural patterns would be like...
Syntactic knowledge but not lexical	Consistent with the co-occurrence frequency of the sequences:  Classifiers: $FFD1 = FFD3 < FFD2$  Nouns: <i>Ba</i> $FFD_{count} = FFD_{mass}$  <i>Gen/kuai/pian</i> $FFD_{count} > FFD_{mass}$
Lexical knowledge but not syntactic	(This is not possible due to the observations that in the Visual World Paradigm Experiment, L2ers have already showed that they were sensitive to the syntactic mass/count cues.)
Both of them	Native-like patterns
None of them	(This is not possible since the Visual World Paradigm Experiment has already showed that L2ers have already acquired the syntactic knowledge of classifiers.)

## 5. Results

In order to get a clear picture about how native Mandarin speakers and late Dutch-Mandarin learners process nominal phrases on-line, and how the mass/count syntactic cues and mass/count-bias properties of classifiers interact, participants' eye movements on the nominal phrases and the specific characters (the classifiers and the nouns) were analysed. The nominal phrases excluding the numeral (i.e. the construction [CL-Adj-N]

in T1, the construction [Adj-CL-N] in T2, and the construction [Adj-CL-*de*-N] in T3), the classifiers, and the nouns were chosen as the Regions of Interest (ROI). To explore both the early and late stages of on-line processing, first fixation durations and total reading times were used as the measures. First Fixation Duration (FFD) is the duration of the first first-pass fixation in a ROI, reflecting participants' early stage of processing and the predictability of each item. Total Reading Time (TRT) is the sum of the durations of all fixations in a ROI, reflecting participants' late stage of processing and the difficulty of understanding the items in that region.

The linear mixed effects modelling approach (LME; Baayen, 2008; Baayen, Davidson, & Bates, 2008) was used to examine participants' on-line processing. By adding random effects from participants and items in the model, the LME approach is more powerful than the traditional ANOVA approach with separate participant (F1) and item (F2) analyses. In addition, the LME approach does not require prior averaging across participants and items; instead, it works with trial-based data set directly. It can handle missing data due to target skipping or tracking errors. The statistical procedure was conducted using the *lmer* program (lme4 package; Bates, Maechler, & Dai, 2008) in the R system (R Development Core Team, 2008).

When using the LME, there is no consensus on which is the best way to do the data analysis: confirmatory or exploratory. Based on the results of the Visual World Paradigm experiment, the main concern of this Reading for Comprehension experiment is to test the effect of nominal phrase structures on L1ers' and L2ers' on-line processing, and also to explore the possible interactions between the mass/count structures and the mass/count-bias properties of different classifiers. Thus the exploratory LME was used. In order to explore the best fit model, for each ROI (i.e. the phrases, the classifiers and the nouns), I started with a 'full model' (Zuur, Ieno, Walker, Saveliev, & Smith, 2009; Barr, Levy, Scheepers, & Tily, 2013) with all the possible factors: the nominal phrase structures (3 levels), the mass/count nouns (2 levels), and the classifiers (4 levels) as the three fixed factors, participants, items, and the co-occurrence frequency of the whole sequence as the three random factors. Based on this 'full model', using the step-wise regression, I reduced the model by systematically removing non-significant terms, and eventually got the best fit model (Yan, Zhou, Shu, Yusupu, Miao, Krugel, & Kliegl, 2014). In the following part of this section, the results of the explored best fit model are presented for each ROI and for both native Mandarin speakers (L1ers) and late Dutch-

Mandarin learners (L2ers), reporting the estimate (*Estimate*), standard errors (*SE*), *t* values (*t*) and the corresponding *p* values (*p*). All the figures in the current study were generated with ggplot2 (Wickham, 2009).

### 5.1. Processing of the phrases

Since each nominal phrase contains three to four items (a classifier, an adjective, a noun, and a modification marker *de* (only in T3)), it is meaningless to analyse FFD in the region of the nominal phrases since the first fixation would always be landed on the first character of the nominal phrase. In this case, only TRT were analysed in this ROI.

All the data beyond mean  $\pm 3$ \* standard deviation (sd) were excluded (Rayner et al., 2010). The rates of outliers' removal of TRT for L1ers and L2ers are 0.91% and 1.09% respectively.

By using the step-wise regression, the best fit models for L1ers and L2ers on TRT in the ROI of the nominal phrases are illustrated in (40).

(40). L1ers: TRT ~ Nominal types \* Classifier + (1|item)

L2ers: TRT ~ Nominal types + Classifier + (1|id) + (1|item)

From the best fit models in (40) we can see that the mass/count status of nouns is not included for either L1ers or L2ers, indicating that the mass/count nouns did not affect both groups' TRT on the nominal phrases. For L1ers, the best fit model includes the nominal types, different classifiers, and their interactions as the fixed factors. To further look into how different nominal phrase types (i.e. different mass/count syntactic cues) affected L1ers' TRT on the nominal phrases with different classifiers, a full contrast among the three nominal structure types with each classifier was done. The results of the interactions between the nominal structure types and different classifiers are summarized in Table 33. Detailed results are illustrated in Appendix B.4.

Table 33. TRT on the nominal phrase for L1ers

<i>Diff</i>	T2-T1	T3-T1	T3-T2
<b><i>Ba</i></b>	138.37	205.97*	66.78
<b><i>Gen</i></b>	76.87	403.47**	326.6*
<b><i>Kuai</i></b>	-69.55	239.13*	153.25*
<b><i>Pian</i></b>	-52.43	187.97*	239.16*

\*\*  $p < 0.01$ , \*  $p < 0.05$

As for the L2ers, both nominal phrase types and different classifiers had an effect on L2ers' TRT, while no interaction between these two factors is included in the best fit model. Also, compared to the best fit model for L1ers, the best fit model for L2ers includes the random effect from participants (id), which means compared to native Mandarin speakers, late Dutch-Mandarin learners exhibited larger individual variations in their TRT on the nominal phrases. The detailed results of the best fit LME model for L2ers are summarized in Table 34.

Table 34. TRT on the nominal phrase for L2ers

	Estimate	Std. Error	df	t-value	p-value
Nominal t1- t2	307	216	118	1.42	0.158
Nominal t1- t3	587	215	115	2.73	0.007**
Nominal t2 - t3	280	215	116	1.3	0.196
Classifier ba - gen	1005	260	117	3.87	2E-04***
Classifier ba - kuai	581	263	119	2.21	0.029*
Classifier ba - pian	834	270	118	3.08	0.002**
Classifier gen - kuai	-425	247	119	-1.72	0.089
Classifier gen - pian	-171	274	117	-0.62	0.534
Classifier kuai - pian	254	274	118	0.92	0.357

\*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$

From Table 34 we can see that nominal phrases in T3 had longer TRT than in T1. Nominal phrases with the classifier *ba* had shorter TRT than nominal phrases with the classifiers *gen/kuai/pian*. The differences of TRT among nominal phrases with the three classifiers *gen*, *kuai*, and *pian* were not significant. The interaction between nominal structure types and classifiers is not included in the best fit LME model, indicating that for each classifier, the nominal structure types affected L2ers' TRT on the phrases in the same way.

## 5.2. Processing of the classifiers

In the ROI of the classifiers, two eye tracking indexes were analysed: FFD and TRT.

### 5.2.1. The early stage of processing – first fixation duration

Previous studies found that in Mandarin reading, fixations shorter than 60ms are not related to the brain's cognitive processing. Participants cannot obtain meaningful and useful information from these short fixations (Wang, Chen, Yang, & Mo, 2008; Yang, Wang, Tong, & Rayner, 2012, etc.). Based on previous studies, all first fixations shorter than 60ms were excluded. Also, the FFD beyond  $\text{mean} \pm 3 * \text{sd}$  were deleted. The rates of outliers' removal of FFD for L1ers and L2ers are 0.61% and 0.79% respectively.

Using the step-wise regression, the best fit models for L1ers and L2ers on the FFD in the ROI of the classifiers are illustrated in (41).

(41). L1ers:  $\text{FFD} \sim \text{Nominal types} * \text{Classifier} + (1|\text{item})$

L2ers:  $\text{FFD} \sim \text{Nominal types} * \text{Classifier} + (1|\text{item}) + (1|\text{id})$

From the best fit models in (37) we can see that for both groups of participants, the mass/count status of nouns is not included in the best fit model, indicating that the mass/count nouns did not affect their early processing of the classifiers, which is reasonable since participants only reach the nouns after they finish processing the classifiers. The best fit model includes the nominal types, different classifiers, and their interactions as the fixed factors. To further look into the effect of different nominal structure types on participants' FFD in classifiers, a full contrast among the three nominal structure types with each classifier was done. The results of the interactions between the nominal structure types and different classifiers are summarized in Table 35, and visualized in Figure 16. Detailed results are illustrated in Appendix B.5.

Table 35. The differences of FFD on the classifier for L1ers and L2ers

<i>Diff</i>	<b>L1ers</b>			<b>L2ers</b>		
	T2-T1	T3-T1	T3-T2	T2-T1	T3-T1	T3-T2
<b><i>Ba</i></b>	67.67*	11.77	-55.91	273.23**	-54.76	-328.01**
<b><i>Gen</i></b>	5.88	24.56	19.75	180.7*	104.8	-75.9
<b><i>Kuai</i></b>	-6.39	19.63	25.55	401.6**	128.5	-273.1*
<b><i>Pian</i></b>	11.16	-2.48	-14.47	111.43	-46.75	-158

\*\*  $p < 0.01$ , \*  $p < 0.05$



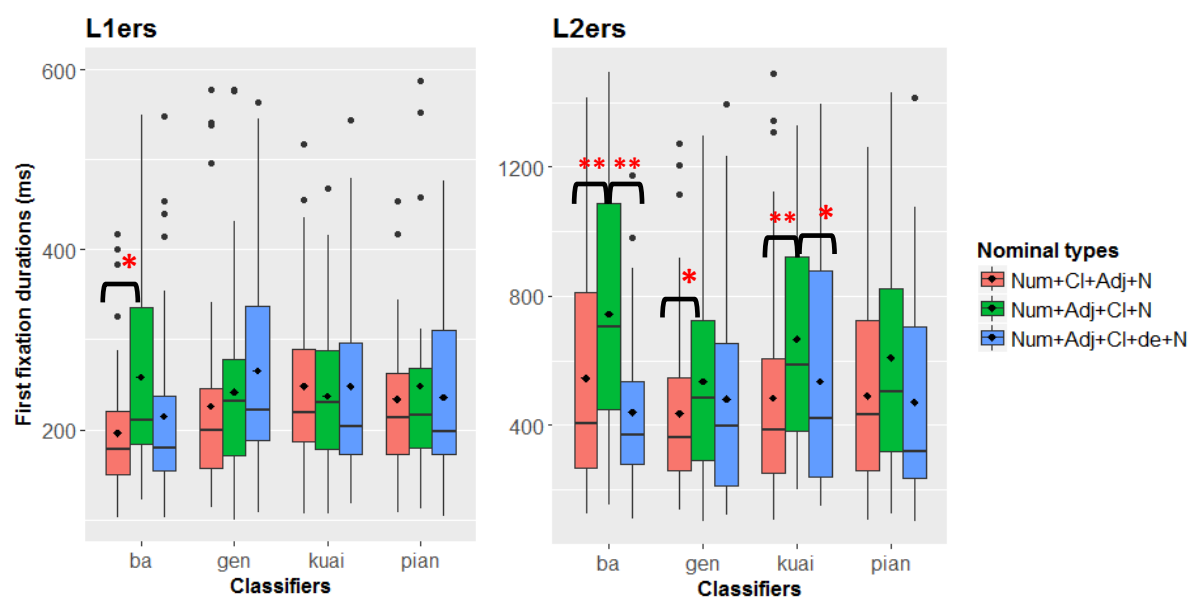


Figure 16. First fixation duration on classifiers

(Boxes with different colours represent the middle 50% range of processing times of the classifier in different nominal types (the red box represents the nominal structure [Num+Cl+Adj+N], the green box represents the nominal structure [Num+Adj+Cl+N], and the blue box represents the nominal structure [Num+Adj+Cl+de+N]). The horizontal line within each box represents the mid-point of processing times in each condition, while the black dot in each box represents the mean value of processing times in each condition. The upper and lower whiskers represent the data of processing times which outside the middle 50% range. \*\*  $p < 0.01$ , \*  $p < 0.05$ )

From Table 35 and Figure 16 we can see that for both L1ers and L2ers, the nominal phrase structures affected their FFD differently with different classifiers. For L1ers, when the classifier was *ba*, the classifier in T1 had significantly shorter FFD than in T2. When the classifier was *gen/kuai/pian*, the differences between any two of the three nominal types were not significant. For L2ers, when the classifier was *ba/kuai*, FFD on the classifier in T2 was significantly longer than in T1 and T3. When the classifier was *gen*, FFD on the classifier in T2 was significantly longer than in T1. When the classifier was *pian*, there was no significant difference among the three nominal types.

### 5.2.2. The late stage of processing – total reading time

All the data beyond  $\text{mean} \pm 3 \times \text{sd}$  were excluded, the rates of outliers' removal of TRT for L1ers and L2ers are 0.89% and 1.15% respectively.

L1ers and L2ers have the same best fit model of TRT in the ROI of the classifiers, which is illustrated in (42).

(42).  $TRT \sim \text{Nominal types} + (1|\text{item})$

From the model in (42) we can see that only nominal phrase structures affected L1ers' and L2ers' TRT in the ROI of the classifiers. For L1ers, TRT on the classifier in T2 was significantly longer than in T1 ( $Estimate = 66.78, SE = 28.55, t = 2.339, p < 0.05$ ). The difference between T1 and T3, and T2 and T3 were not significant ( $Estimate = 42.65, SE = 28.56, t = 1.494, p = 0.138$ ;  $Estimate = -24.15, SE = 28.35, t = -0.851, p = 0.396$ ). For L2ers, TRT on the classifiers in T2 was significantly longer than in T1 ( $Estimate = 348.5, SE = 85.81, t = 4.061, p < 0.001$ ) and in T3 ( $Estimate = 320.61, SE = 86.25, t = 3.727, p < 0.001$ ). The difference between T1 and T3 was not significant ( $Estimate = 27.87, SE = 86.36, t = 0.323, p = 0.747$ ). Results are visualized in Figure 17.

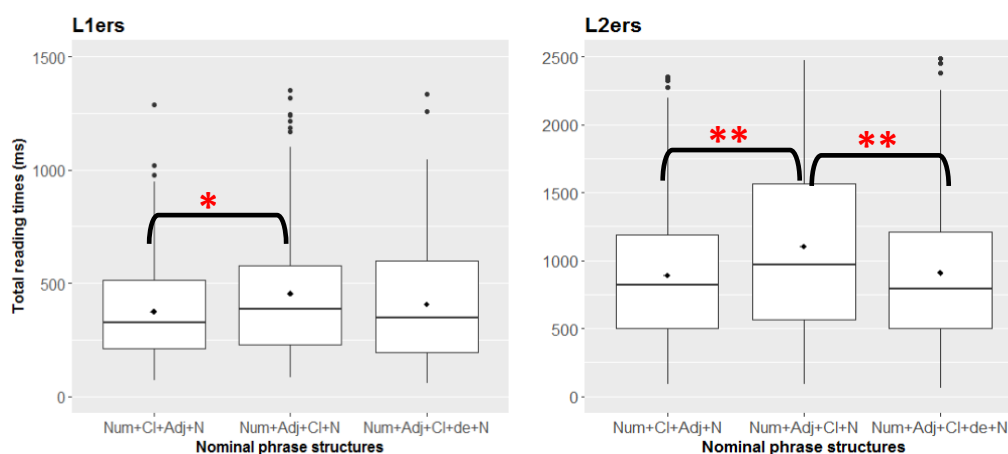


Figure 17. Total reading times on classifiers

### 5.3. Processing of the nouns

In the ROI of the nouns, two eye tracking index were analysed: FFD and TRT.

#### 5.3.1. The early stage of processing – first fixation duration

All first fixations shorter than 60ms were deleted. Besides, FFD beyond  $\text{mean} \pm 3 * \text{sd}$  were deleted. The rates of outliers' removal of FFD for L1ers and L2ers are 0.74% and 0.98% respectively.

According to the step-wise regression, the best fit models for L1ers and L2ers of FFD in the ROI of the nouns are illustrated in (43).

(43). L1ers: FFD ~ Nominal types \* Noun \* Classifier + (1|item)

L2ers: FFD ~ Noun \* Classifier + (1|id) + (1|item)

From the models in (43) we can see that, L1ers and L2ers have different models. For L1ers, apart from the nominal phrase structures and the mass/count nouns, different classifiers had an effect on their FFD on the nouns. Also, there are some interactions among these three factors. For L2ers, different from L1ers, only the mass/count nouns, different classifiers, and their interactions affected L2ers' FFD on the nouns, but not different nominal types.

In order to further look into how different nominal phrase structures and mass/count nouns affected L1ers' FFD on the nouns with different classifiers, exploratory LME models were run separately for different classifiers.

#### 5.3.1.1. When the classifier was *ba*

When the classifier was *ba*, the best fit model for L1ers is illustrated in (44).

(44). FFD ~ Nominal type \* Noun + (1|item)

From the model in (44) we can see that the different nominal types affected L1ers' FFD on the nouns differently with count or mass nouns. To breakdown the interactions between nominal types and the mass/count nouns, a full contrast was done. The results are summarized in Table 36 and visualized in Figure 18. Detailed results are illustrated in Appendix B.6.

Table 36. The differences of FFD on nouns with *ba* for L1ers

	Count Nouns			Mass Nouns		
	T2-T1	T3-T1	T3-T2	T2-T1	T3-T1	T3-T2
<i>Diff</i>	40.96*	21.22*	-19.74	-38.76*	-80.41*	-41.64

\*  $p < 0.01$

From Table 36 and Figure 18 we can see that count nouns and mass nouns had different FFD patterns among nominal structure types when the classifier was *ba*. When the nouns were count nouns, FFD in T2 and T3 was significantly longer than in T1 (*Estimate* = 40.96, *SE* = 19.21,  $t = 2.132$ ,  $p < 0.05$ ; *Estimate* = 21.22, *SE* = 11.94,  $t = 1.864$ ,  $p < 0.05$ ). When the nouns were mass nouns, FFD in T2 and T3 was significantly shorter than in

T1 (*Estimate* = 38.76, *SE* = 21.46, *t* = 2.68, *p* < 0.05; *Estimate* = 80.41, *SE* = 27.46, *t* = 2.979, *p* < 0.05).

#### 5.3.1.2. When the classifier was *gen*

When the classifier was *gen*, the best fit model for L1ers is illustrated in (45).

(45). FFD ~ Noun + (1|item)

From the model in (45) we can see that only the mass/count nouns affected L1ers' FFD on the nouns when the classifier was *gen*. The results of the best fit model showed that count nouns had marginally significantly longer FFD than mass nouns when the classifier was *gen* (*Estimate* = 21.8, *SE* = 14.22, *t* = 1.49, *p* = 0.061).

#### 5.3.1.3. When the classifier was *kuai*

When the classifier was *kuai*, the best fit model for L1ers is illustrated in (46).

(46). FFD ~ Noun + (1|item)

The results of the best fit model showed that count nouns had significantly longer FFD than mass nouns when the classifier was *kuai* (*Estimate* = 38.7, *SE* = 15.8, *t* = 2.45, *p* < 0.05).

#### 5.3.1.4. When the classifier was *pian*

When the classifier was *pian*, the best fit model for L1ers is illustrated in (47).

(47). FFD ~ Noun + (1|item)

The results of the best fit model showed that count nouns had significantly longer FFD than mass nouns when the classifier was *pian* (*Estimate* = 41.7, *SE* = 16.8, *t* = 2.49, *p* < 0.05).

In general, for L1ers, the patterns of FFD on the nouns varied based on different classifiers. The patterns of FFD on both count nouns and mass nouns across all the four classifiers are summarized in Table 37 and visualized in Figure 18. FFD1 stands for FFD in T1, and so on.

Table 37. The FFD patterns in the ROI of the nouns for L1ers

Classifier	Noun	Patterns of first fixation durations
<i>Ba</i>	Count	FFD1 < FFD2 = FFD3
	Mass	FFD1 > FFD2 = FFD3
<i>Gen</i>	Count/Mass	FFD1 = FFD2 = FFD3
<i>Kuai</i>	Count/Mass	FFD1 = FFD2 = FFD3
<i>Pian</i>	Count/Mass	FFD1 = FFD2 = FFD3

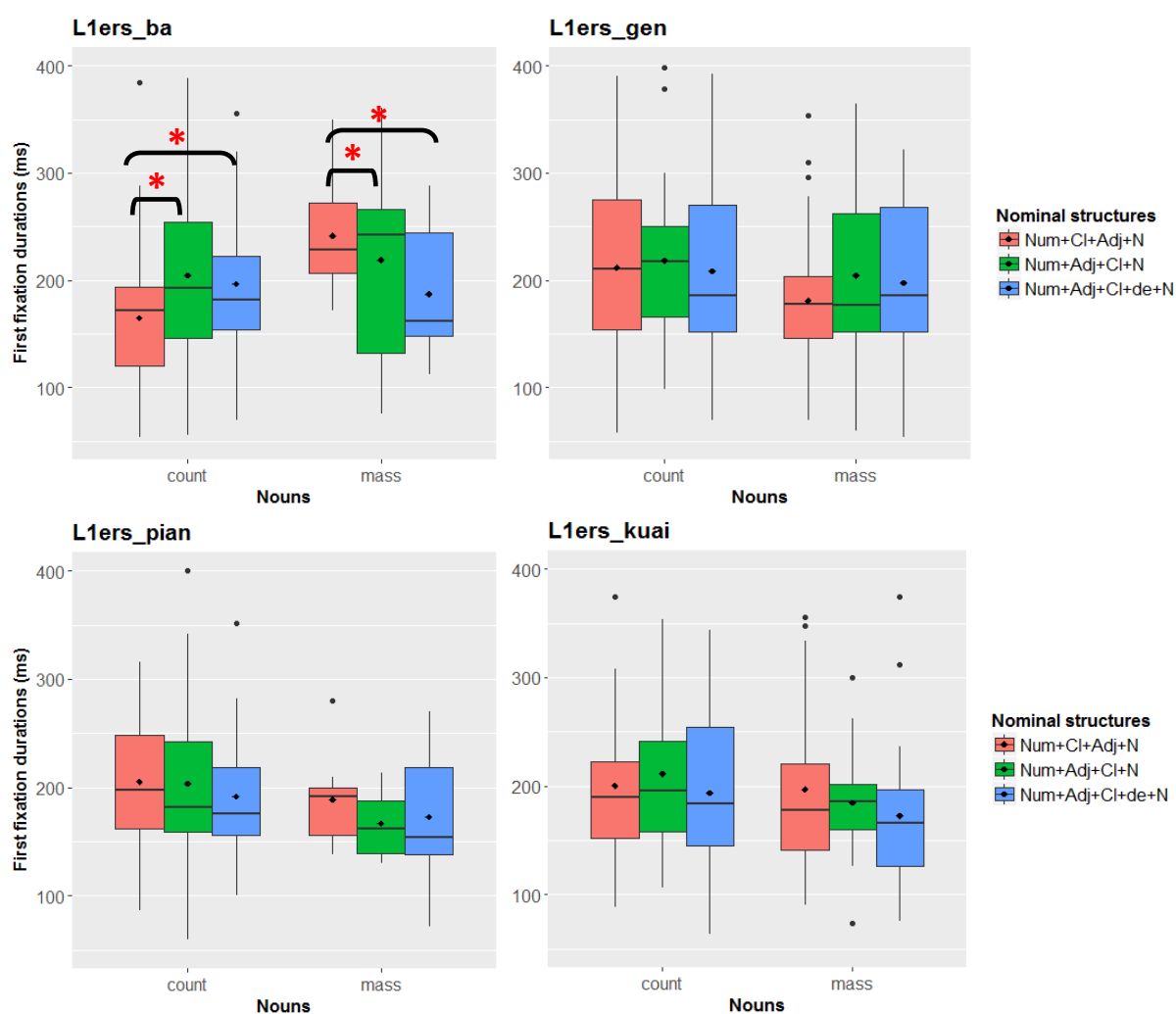


Figure 18. First fixation duration on nouns with different classifiers for L1ers

Regarding L2ers, a full contrast was done to further look into the effect of mass/count nouns on their FFD in the ROI of the nouns with different classifiers. The results are summarized in Table 38, and detailed results are illustrated in Appendix B.7.

Table 38. The difference of FFD between count and mass nouns with different classifiers for L2ers

<i>Diff</i>	<i>Ba</i>	<i>Gen</i>	<i>Kuai</i>	<i>Pian</i>
Mass-Count	-1.4	-153.4*	-98.7.	-278.1**

\*\*  $p < 0.01$ , \*  $p < 0.5$ , .  $p = 0.062$

Both L1ers' and L2ers' FFD on mass/count nouns with different classifiers are visualized in Figure 19.

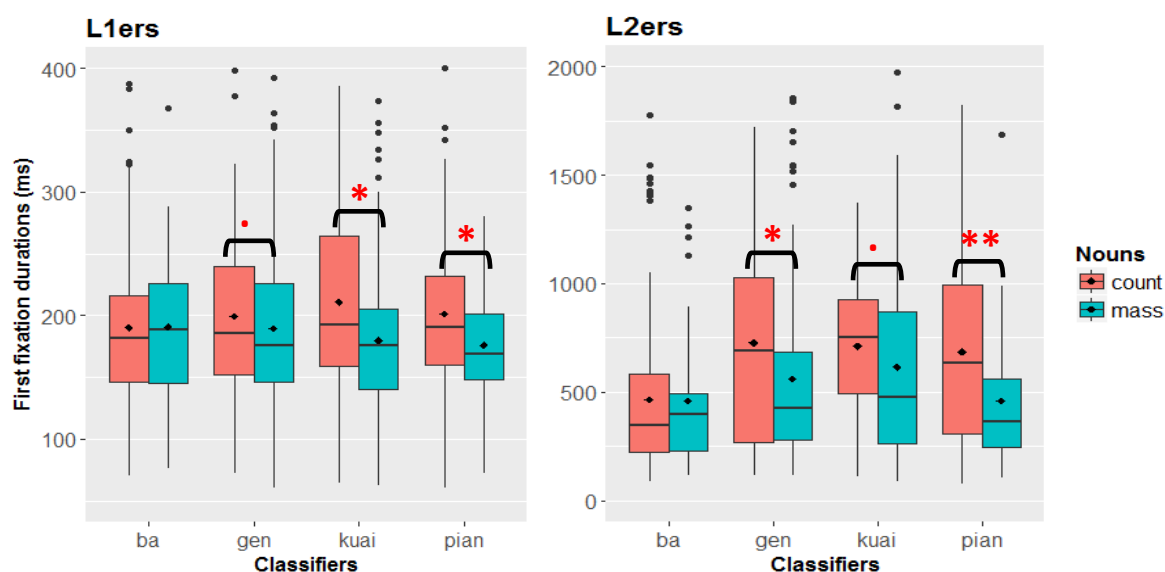


Figure 19. First fixation durations on the nouns

From Figure 19 we can see that L1ers and L2ers had similar patterns on FFD in the ROI of the nouns. When the classifier was *ba*, the difference between count nouns and mass nouns was not significant. When the classifier was *gen/pian/kuai*, FFD on count nouns was longer than it on mass nouns (significantly or marginally significantly).

### 5.3.2. The late stage of processing – total reading time

All the data beyond  $\text{mean} \pm 3 \times \text{sd}$  were deleted, the rates of outliers' removal of TRT for L1ers and L2ers are 0.95% and 1.21% respectively.

L1ers and L2ers have the same best fit model on the TRT in the ROI of the nouns based on the step-wise regression, which is illustrated in (48).

(48).  $\text{TRT} \sim \text{Classifier} + (1|\text{item})$

From this model we can see that for both L1ers and L2ers, neither the nominal types nor the mass/count nouns had an influence on TRT in the ROI of the nouns. It is the different classifiers which affected their late stage processing of the nouns. A full contrast was done to compare the differences between any two of the four classifiers. The results are summarized in Table 39 and visualized in Figure 20. Detailed results are illustrated in Appendix B.8.

Table 39. The differences of TRT on the nouns

<i>Diff</i>	<b>L1ers</b>	<b>L2ers</b>
<i>Ba vs. Gen</i>	128.24*	619***
<i>Ba vs. Kuai</i>	84.34*	353*
<i>Ba vs. Pian</i>	143.38*	511***
<i>Gen vs. Kuai</i>	-44.16	-266
<i>Gen vs. Pian</i>	15.51	-108
<i>Kuai vs. Pian</i>	59.36	246

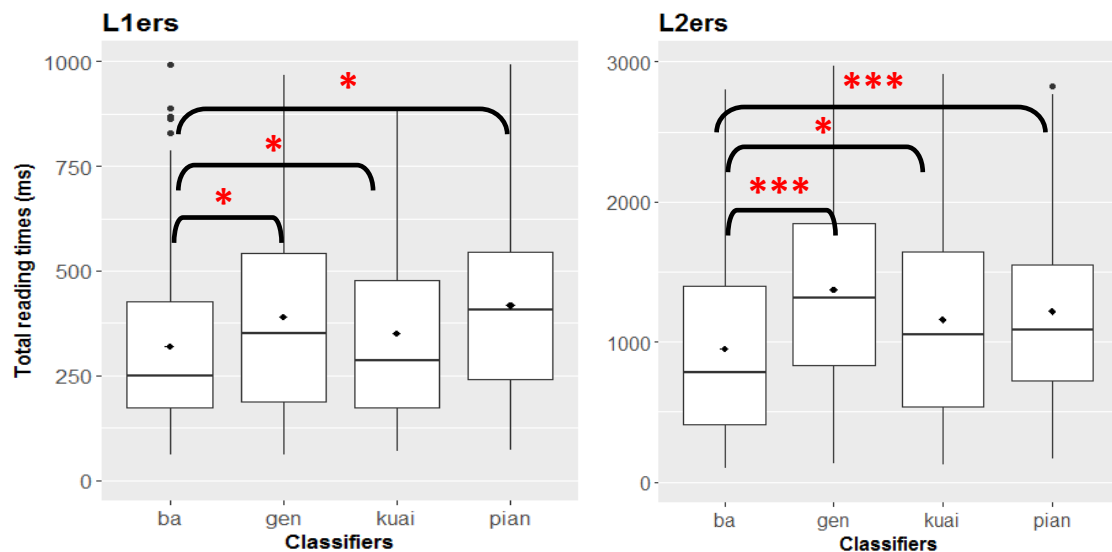


Figure 20. Total reading times on the nouns

## 6. Discussion

In current study, the eye-tracking technique and the Reading for Comprehension task were used to look into the effects of different nominal phrase structures and different classifiers on native Mandarin speakers' and late Dutch-Mandarin learners' on-line processing. The four dual-role classifiers used in the current study were *ba* 'handful', *gen* 'rod', *pian* 'slice', and *kuai* 'chunk', all of which are ambiguous words since they all possess both count-classifier meanings and massifier meanings. Nominal phrase structures were manipulated to have three types: nominal phrases in T1 (Type 1) has the structure [Num+CL+Adj+N], in T2 has the structure [Num+Adj+CL+N], and in T3 has the structure [Num+Adj+CL+*de*+N]. For each classifier, associated nouns were chosen based on their typicality as mass/count nouns (based on the results in the Noun Rating Test). All the sentence materials are grammatical and acceptable to native Mandarin speakers (based on the results in the Grammaticality Judgement Test).

According to Cheng & Sybesma (1998, 1999), Borer (2005), and Li et al (2008), different nominal phrase structures could lead readers to have different interpretations for the phrases: when there is an adjective preceding the classifier, and/or a modification marker *de* inserted after the classifier, the nominal phrase should be interpreted as a mass or a plural meaning. Also the Visual World Paradigm experiment and the Cloze Probability Test reported in Chapter 4 showed that different classifiers have different ambiguity types and classifier-noun restrictions. Based on the results of the Cloze Probability Test, the classifier *ba* is a count-biased classifier which has a dominant count-classifier meaning and strong connections with associated count nouns, while the classifiers *gen/kuai/pian* are mass-/equally-biased classifiers which do not have any preference for count nouns or mass nouns. Putting a count-biased classifier in a mass-preferred structure may cause some processing difficulties on participants. In the Reading for Comprehension experiment, longer fixation times on an item could be treated as an index of processing difficulty or unexpected items. Hence, I predict that the classifier *ba* in the mass-biased structures (T2 and T3) will have longer fixation durations than in the neutral structure (T1), while mass-biased classifiers will have similar fixations times among all three types of nominal structures, and equally-biased classifier will have shorter fixation durations in T2 and T3 than in T1 since the mass-biased structure exclude the competition between the two equally-possible meanings.



To have a full understanding of both L1ers' and L2ers' on-line processing and interpretation of nominal phrases, the nominal phrases, the classifiers and the nouns were set as the three ROI. LME and the step-wise regression were used to explore the best fit model for both FFD and TRT. Since every nominal phrase includes a classifier and a noun, the processing times of each item will have an influence on the processing of the nominal phrase. In the following part of this section, the classifiers' results are discussed first, followed by the nouns'. The TRT of the nominal phrases are discussed in the end. Before discussing both L1ers' and L2ers' processing times on the classifiers, the nouns, and the phrases, the division among classifiers are discussed.

### 6.1. Classifiers' division on L1ers and L2ers

From the results reported above we can see a clear effect of the classifiers on participants' on-line processing. This effect is observed in the early stage processing of the classifiers and the nouns, as well as the late stage processing of the phrases. To be specific, for L1ers, the best fit LME models for FFD in the ROI of the classifiers and the nouns, and the best fit LME models for TRT in the ROI of the phrases all include 'classifier' as a fixed factor. Also the interactions between nominal phrase types and different classifiers are significant. In other words, for L1ers, the three types of nominal phrase structures affected FFD in the ROI of the classifiers and the nouns, and TRT in the ROI of the phrases differently with different classifiers. The patterns of processing times on the classifiers, the nouns and the nominal phrases with different classifiers are summarized in Table 40. FFD is short for First Fixation Durations, and TRT is short for Total Reading Times. The numerals 1, 2, and 3 represents the three nominal phrase types T1, T2, and T3 respectively. FFD1 stands for First Fixation Durations in T1, and so on. FFD<sub>count</sub> stands for First Fixation Durations on the count nouns while FFD<sub>mass</sub> stands for First Fixation Durations on the mass nouns.

Table 40. The pattern of processing times in different ROI with different classifiers for L1ers

Classifiers		FFD on classifiers	FFD on nouns	TRT on phrases
<i>Ba</i>	Count N	FFD1 = FFD3 < FFD2	FFD1 < FFD2 = FFD3	TRT1 < TRT2 = TRT3
	Mass N		FFD1 > FFD2 = FFD3	
<i>Gen</i>		FFD1 = FFD2 = FFD3	FFD <sub>count</sub> > FFD <sub>mass</sub>	TRT1 = TRT2 < TRT3
<i>Kuai</i>		FFD1 = FFD2 = FFD3	FFD <sub>count</sub> > FFD <sub>mass</sub>	TRT1 = TRT2 < TRT3
<i>Pian</i>		FFD1 = FFD2 = FFD3	FFD <sub>count</sub> > FFD <sub>mass</sub>	TRT1 = TRT2 < TRT3

From Table 40 we can see that in all the three ROIs, L1ers always exhibited different behaviour patterns when the classifier was *ba* compared to when the classifier was *gen/kuai/pian*. Based on L1ers' behaviours, the four classifiers can be divided into two groups: *ba* by itself is in one group, while *gen*, *kuai*, and *pian* are in another. The particularity of the classifier *ba* is observed not only in the early stage of processing (reflected by FFD on the classifiers and the nouns), but also in the late stage of processing (reflected by TRT on the nominal phrases). This is consistent with the results of the Visual World Paradigm experiment and the revised Cloze Probability Test. As discussed in Chapter 4, the four classifiers have different properties (massifier meanings, ambiguity types, preferred mass/count nouns, co-occurrence frequency of nouns). These differences are summarized in Table 41.

Table 41. The differences among the four classifiers

	Massifier meaning	Ambiguity type	Preferred nouns	Co-occurrence frequency
<b><i>Ba</i></b>	Collective	Count-biased	Count nouns	Count = Mass
<b><i>Gen</i></b>	Divider	Mass/Equal-biased	Count/Mass nouns	Count < Mass
<b><i>Kuai</i></b>	Divider	Mass/Equal-biased	Count/Mass nouns	Count < Mass
<b><i>Pian</i></b>	Divider	Mass/Equal-biased	Count/Mass nouns	Count < Mass

From Table 41 we can see that the division of the four classifiers based on L1ers' eye movement performances is consistent with the divisions based on their massifier meanings, ambiguity types, preferred nouns' count/mass categories, and the co-occurrence frequency with count and mass nouns. This finding indicates that classifiers' properties have some influences on native Mandarin speakers' on-line processing.

Different from L1ers, L2ers did not exhibit the consistent classifier division pattern during their on-line processing. Their processing patterns are summarized in Table 42.

Table 42. The patterns of processing times in different ROI with different classifiers for L2ers

Classifiers	FFD on classifiers	FFD on nouns	TRT on phrases
<i>Ba</i>	FFD1 = FFD3 < FFD2	FFD <sub>count</sub> = FFD <sub>mass</sub>	TRT1 = TRT2 < TRT3
<i>Gen</i>	FFD1 < FFD2	FFD <sub>count</sub> > FFD <sub>mass</sub>	TRT1 = TRT2 < TRT3
<i>Kuai</i>	FFD1 = FFD3 < FFD2	FFD <sub>count</sub> > FFD <sub>mass</sub>	TRT1 = TRT2 < TRT3
<i>Pian</i>	FFD1 = FFD2 = FFD3	FFD <sub>count</sub> > FFD <sub>mass</sub>	TRT1 = TRT2 < TRT3

From Table 42 we can see that for L2ers, only the FFD pattern in the ROI of the nouns exhibited the division between the classifier *ba* and the other three classifiers, but not the FFD patterns in the ROI of the classifiers and TRT patterns in the ROI of the phrases. Based on these results we could say that, even having reached an advanced level of Mandarin proficiency, late Dutch-Mandarin learners had not implicitly acquired the count/mass properties of different classifiers (the ambiguity type, the count/mass preference of associated nouns). They did not exhibit native-like behaviours in real time processing of different classifiers during on-line processing. In the future, the revised Cloze Probability Test needs to be conducted on high proficiency L2ers to explore what is the exact reason of L2ers' non-native-like behaviours on the on-line processing of different classifiers.

## 6.2. The processing of the classifiers

### 6.2.1. The early stage of processing of the classifiers (first fixation durations)

The processing of the classifiers thus can be affected by two factors: their count/mass properties (meanings of being massifiers, ambiguity types, preferred mass/count nouns, and the co-occurrence frequency with count/mass nouns) and the nominal phrase structures. Based on the results from the Cloze Probability Test, the classifier *ba* is a count-biased classifier which has a dominant count-classifier reading, while the classifiers *gen/kuai/pian* are either mass-biased classifiers or equal-biased classifiers. According to previous studies (Rayner & Duffy, 1986, etc.), different types of ambiguous words would behave differently in different nominal structure types. The predictions of FFD on different types of classifiers in different nominal phrase structures are summarized in Table 27, repeated here as in Table 43.

Table 43. The predictions of FFD on different classifiers

		Count-biased classifier	Mass-biased classifier	Equally- biased
T1[Num+CL+Adj+N]	Neutral			
T2[Num+Adj+CL+N]	Mass- biased	$T1 < T2 = T3$	$T1 = T2 = T3$	$T1 > T2 = T3$
T3[Num+Adj+CL+ <i>de</i> +N]	Mass- biased			

The best fit LME models for L1ers and L2ers both include nominal phrase structures, classifiers and their interactions as the fixed factors, indicating that for both of the two groups, nominal phrase structures affected their FFD in the ROI of the classifiers

differently with different classifiers. However, the FFD patterns with different classifiers are different between L1ers and L2ers. For L1ers, when the classifier was *ba*, the classifiers in T2 had significantly longer FFD than in T1 and T3. The difference between T1 and T3 was not significant. It is reasonable to observe longer FFD on the classifier *ba* in T2 than in T1 if considering *ba* as a count-biased classifier. According to Cheng & Sybesma (1998, 1999), the Adj-CL order in T2 is a syntactic cue for massifiers. When processing nominal phrases in T2, on reaching the adjective directly following the numeral, L1ers have already built the anticipation that the following item should be a massifier. When reaching *ba* which has a dominant count-classifier meaning, a conflict between the dominant count-classifier meaning from the classifier and the preferred massifier meaning from the structure occurred. On realizing this conflict, participants need to suppress the dominant count-classifier meaning of *ba* and activate the less-likely massifier meaning in the very early stage of on-line processing. This suppression and activation procedure caused some extra processing times, reflected by the longer FFD on the classifier *ba* in T2 than in T1. By the same token, it would be expected to observe longer FFD on *ba* in T3 than in T1 since the nominal phrases in T2 and T3 share the same word order (Adj-CL). However, FFD on the classifier *ba* in T3 had no significant difference with it in T1. For this unexpected finding, there is one possibility – the modification marker *de* in T3. It should be noted that nominal phrases in T2 and T3 have the exact same items and structure till participants reach the modification marker *de* in T3, which follows the classifier. The shorter FFD on the classifier *ba* in T3 than in T2 may be caused by this extra modification marker *de*. In Chinese reading, the perceptual span extends one character to the right of the fixation point (Inhoff & Liu, 1998, Rayner & Duffy, 1986; Yen, Radach, Tzeng, Hung, & Tsai, 2009), and the lexical information of the character next to the current fixed character could have some influence on the current processing ('parafoveal-on-foveal effect', Rayner & Duffy, 1986). Also, the 'parafoveal-on-foveal' effect takes place at the very early stage of on-line processing (around 40-140ms) (Yen, Radach, Tzeng, Huang, & Tsai, 2009). In the current study, when reading the nominal phrases in T3, on reaching the classifier *ba*, participants may already have parafovealled the modification marker *de*. Even though the word order [Adj-CL] force a massifier reading on the classifier *ba*, the parafovealled information of the modification marker *de* would force a measurement reading (the meaning of the amount/volume of the objects, see Cheng & Sybesma (1998), as discussed in Chapter 2) on the phrase. In Mandarin, even though mass-biased classifiers can fit in a nominal

phrase with a measurement reading, it is not necessary for the massifiers to be concrete (i.e. the container/measurement does not necessarily physically exist). However, it is necessary that the massifier in a mass-biased but not measurement structure to be concrete (Similar to the difference between ‘three bottles worth of wine’ vs. ‘three bottles of wine’). In this case, it is reasonable to assume that even *ba* had to be interpreted as a massifier in a measurement structure (T3); it is the semantic feature of the ‘abstract measurement/container’ which facilitated the processing of the classifier *ba* in T3. This possibility could also explain the observation that FFD on *ba* in T3 were marginally significantly shorter than in T2 ( $p = 0.063$ ).

Different from *ba*, when the classifier was *gen/kuai/pian*, classifiers in all three types of nominal phrase structures had similar FFD patterns. The differences of FFD in the ROI of the classifiers among the three nominal types were not significant. As discussed above in Table 37, mass-biased classifiers would have similar processing times among the three types of nominal structures since the dominant massifier meaning could fit in any of the three structures. Thus based on the results of FFD patterns in the ROI of the classifiers, I argue that the classifiers *gen/kuai/pian* are mass-biased classifiers.

Different from L1ers, the best fit LME models for L2ers include participants as a random factor, indicating that compared to L1ers, L2ers’ behaviours were less consistent and had bigger variations. Also, the FFD patterns in the ROI of the classifiers in each nominal phrase structure with different classifiers were different between L1ers and L2ers. For L2ers, when the classifiers were *ba* and *kuai*, FFD on the classifiers in T2 was significantly longer than in T1 and T3. The difference between T1 and T3 was not significant. When the classifier was *gen*, FFD on the classifier in T2 was significantly longer than in T1. The differences between T1 and T3, and T2 and T3 were not significant. When the classifier was *pian*, there was no significant difference among the three nominal types. These findings firstly indicate that different from L1ers, L2ers were not sensitive to the difference between different types of ambiguous classifiers. They had not acquired the implicit knowledge that the classifier *ba* is a count-biased classifier while the classifiers *gen/kuai* are mass-biased classifiers. Also, considering the fact that L2ers could only acquire this knowledge by implicit learning through daily language exposure and experience, the distributional properties of different classifiers and the classifier-noun pairs are expected to have a strong influence on them (Inhoof & Rayner, 1986; Rayner & Duffy, 1986). By analysing the distributional data from two different

Mandarin corpora (National Language Resources Monitoring and Research Centre, Broadcast Media Language Branch, 2009; Ministry of Education and Institute of Applied Linguistics, 2009) it is revealed that the four classifiers used in the current study share the same pattern of the co-occurrence frequency of the sequences: the sequences in the structure [Num+CL+Adj] (T1) and [Num+Adj+CL+*de*] (T3) have higher frequency than the sequences in the structure [Num+Adj+CL] (T2),  $ps < 0.01$ ; the difference between the sequences in T1 and T3 is not significant. In this case, the FFD patterns of the classifier *ba* and *kuai* can be explained by the co-occurrence frequency of the sequences: FFD on the classifiers in T1 had no significant difference with T3, but both were significantly shorter than in T2, which is consistent with the co-occurrence frequency pattern. In addition, the classifier *pian*'s pattern can be explained by its special distributional properties. Different from the classifiers *ba/gen/kuai* which are always associated with concrete nouns, the classifier *pian* more often combines with abstract nouns (e.g. blank, darkness, silence, depression, wailing) than concrete nouns. It is reasonable to assume that L2ers treated the classifier *pian* as a mass-biased classifier since in their experience *pian* more often occurs with the abstract mass nouns. And massifiers would have similar processing times among the three types of nominal phrase structures as predicted in Table 37. Hence it is reasonable to observe that L2ers' FFD on the classifier *pian* had no significant difference among the three nominal types.

Taken together, from the early stage processing of the classifier we can see that, L1ers were sensitive to the mass/count structures (the CL-Adj vs. Adj-CL word order, and the insertion of *de*), and different classifiers' properties. Apart from the Adj-CL word order, the insertion of the modification marker *de* also had an effect on L1ers' on-line processing. Different from L1ers, L2ers were not sensitive to the properties of different classifiers. The distributional properties had a strong influence on L2ers.

### 6.2.2. The late stage of processing of the classifiers (total reading times)

The best fit LME models for the two groups of participants' TRT in the ROI of the classifiers are the same: both have the nominal phrase structures as the only fixed factor. L1ers and L2ers exhibited similar TRT patterns in the ROI of the classifiers: the classifiers in T2 had the longest TRT among the three nominal types, while the difference between T1 and T3 was not significant. This pattern is consistent with the co-occurrence frequency of the sequences: the sequences in T1 and T3 have higher frequency than the sequences in T2,  $ps < 0.01$ . It is reasonable to observe that the classifiers in the nominal

phrases with higher frequency (in T1 and T3) exhibited shorter processing times than in the nominal phrases with lower frequency (in T2).

In general, the TRT patterns in the ROI of the classifiers for both L1ers and L2ers indicate that the late stage processing of the classifiers was influenced by the co-occurrence frequency of the sequences. Nominal phrases with higher frequency needed shorter processing times than the nominal phrases with lower frequency. The properties of different classifiers did not affect participants' late stage processing of classifiers.

### **6.3. The processing of the nouns**

#### ***6.3.1. The early stage processing of the nouns (first fixation durations)***

Previous studies demonstrated that processing a word can be influenced by its preceding context. Readers are usually faster and more accurate in processing words that are congruent with their preceding context or expected based on preceding context (Duffy, Henderson, & Morris, 1989; Stanovich & West, 1981). Some eye-tracking studies found that fixation durations and gaze durations are usually shorter for highly expected words than unexpected words (Dambacher, Goellner, Nuthmann, Jacobs, & Kliegl, 2008; Kliegl, Grabner, Rolfs, & Engbert, 2004; Rayner, Ashby, Pollatsek, & Reichle, 2004).

As discussed in Chapter 2, a classifier is obligatory when counting nouns in Mandarin Chinese. Some researchers argued that there is a selective semantic relationship between a classifier and its associated nouns. Ahrens (1995) and Zhang (2007) argued that each classifier is associated with a certain class of nouns by agreeing on some type of semantic features (e.g. animacy, shape, size, minimal unit). On reading the classifier, this specific classifier-noun relation would force participants to prefer its associated nouns to be the upcoming noun, and thus facilitate the processing of the upcoming noun if it is associated with the classifier. The semantic classifier-noun relationship in Mandarin Chinese has been found in several studies. Chu (2015) examined the influence of classifiers on the activation of compounds. He used a Grammaticality Judgement task to test participants' reaction to the mismatches between the classifiers and the compounds. There were two conditions in his study. In the experimental condition, there were three types of combinations of classifiers and compound words: 1) the classifier matched only with the first constituent of the compound word (which is not the head of the compound noun. Usually the head of the compound noun is the second constituent of it); 2) the classifier matched with the compound word (the target); and 3) the classifier matched with neither

the first constituent nor the compound word. In the control condition, there were three types of nominal phrases: a target nominal, and two other nominal phrases with classifiers neither matched with first constituent nor the compound word. During the experiment, a set of three grammaticality judgment stimuli was displayed simultaneously in the middle of the screen in three lines. Participants were asked to indicate which one of the three nominal phrases is correct by pressing corresponding buttons as accurately and quickly as possible. Participants' response latency and accuracy were recorded. The author found that compared to the control condition, the grammaticality judgment latency was longer in the experimental condition in which the distractor phrases contained a mismatched classifier which is associated with the first constituent of the compound word. The results indicated that the processing of a classifier evokes the activation of its associated nouns. To be specific, the processing of the classifier leads participants to have a strong expectation for its associated nouns to be the upcoming noun. In the experimental condition when the distractor phrase contained a classifier associated with the first constituent of the compound, reaching the first constituent of the compound made participants confirm their expectation and then be ready to finish the analysis of the nominal phrase. Upon reaching the second constituent of the compound, participants started to realize that their analysis was incorrect and they needed to reanalyse the whole nominal phrase. This procedure of realizing mistakes and reanalysing caused the longer response latency in the experimental condition than in the control condition.

The classifier-noun relationship can cause some processing difficulty for readers when the upcoming noun is not associated with the classifier. A significant N400 effect (a typical index for semantic anomaly) has been found in mismatched classifier-noun pairs in several ERP studies. Chou, Huang, Lee, & Lee (2014) conducted an ERP experiment to explore when and how readers make use of the classifier information to predict the succeeding nouns. They manipulated the classifier-noun pairs to have two levels of classifier constraint strength (strong and weak), and three levels of cloze probability for the pairing nouns (high, low, and implausible). They observed an N400 response under both conditions (weak and strong semantic constraint strength) when the succeeding noun was implausible or with low cloze probability of combining with the classifier. The results indicated that readers can use the preceding classifier to predict the upcoming



nouns before the noun's appearance. When the noun is different from what participants expected based on the classifiers, an N400 is evoked.

In addition, Kanero, Imai, Okada, & Hoshino (2015) adopted an ERP experiment to consider whether the mass/count distinction is realized in Japanese grammar. They manipulated the mismatch between nouns and classifiers to have two levels: mismatched nouns and classifiers which are from the same count/mass ontological category (e.g. a count noun combined with a mismatched count-classifier, or a mass noun combined with a mismatched massifier), and mismatched nouns and classifiers which are from different count/mass ontological category (e.g. a count noun combined with a mismatched massifier, or a mass noun combined with a mismatched count-classifier). They found that no matter whether the nouns are count nouns or mass nouns, both levels of the classifier-noun mismatches evoked an N400 response. The results indicated that if a noun is not associated with a classifier, there would be an N400 which signals readers' surprise of reaching an unexpected item, regardless of whether the mismatch is within the ontological category or cross the ontological boundary.

To conclude, it has been shown by several studies that the processing of nouns could be facilitated by the processing of their associated classifiers, and the mismatched classifier-noun pairs would evoke an N400 response which signals processing difficulty. Considering the results from the revised Cloze Probability Test and the FFD patterns on the classifiers (the classifier *ba* is a count-biased classifier which prefers to occur with count nouns, while the classifiers *gen/kuai/pian* are mass-biased classifiers which have no preference for either count or mass nouns), it would be expected to observe shorter fixations on count nouns than mass nouns when the classifier was *ba* and the nominal phrase had a neutral structure (T1), but similar fixation durations between count nouns and mass nouns when the classifier was *gen/kuai/pian* in all three types of nominal phrase structures. The prediction of FFD in the ROI of the nouns with different classifiers and in different nominal phrase structures are summarized in Table 28, repeated here in Table 44.

Table 44. The prediction of FFD on nouns with different classifiers

	Ba		Gen/Kuai/Pian	
	Count N	Mass N	Count N	Mass N
T1[Num+CL+Adj+N]				
T2[Num+Adj+CL+N]	T1<T2=T3	T1>T2=T3	T1=T2=T3	
T3[Num+Adj+CL+ <i>de</i> +N]				

The best fit LME model for L1ers illustrated a three-way interaction among nominal phrase structures, mass/count status of nouns, and different classifiers. Further comparisons showed that when the classifier was *ba*, the interaction between nominal types and mass/count nouns was significant. When the noun was a count noun, FFD on it in T1 was significantly shorter than in T2 and T3. The difference between T2 and T3 was not significant. When the noun was a mass noun, FFD on it in T1 was significantly longer than in T2 and T3. The difference between T2 and T3 was not significant. Different from *ba*, when the classifier was *gen/kuai/pian*, nominal phrase structures did not affect participants' FFD on the nouns. The only effective factor is the mass/count status of the noun. When the classifier was *gen*, FFD on the count nouns was marginally significantly longer than mass nouns ( $p = 0.061$ ). When the classifiers were *kuai* and *pian*, FFD on the count nouns were significantly longer than mass nouns.

As a count-biased classifier, *ba* has strong connections with its associated count nouns, as revealed in the Cloze Probability Test. Also, in the Visual World Paradigm experiment, in the analysis window of the classifier, on hearing *ba*, L1ers started to fix more on the count-noun denoting pictures than the mass-noun denoting pictures. The processing of *ba* is very likely to prime its associated count nouns to be the upcoming noun in a neutral structure (T1). This explains the finding that FFD on count nouns was shorter in T1 than in T2 and T3, since the mass-biased structure (T2 & T3) coerced the count-biased classifier *ba* to be a massifier, which would not facilitate its associated count nouns' processing. This is consistent with the results in the Visual World Paradigm experiment that in Condition 3 where the nominal phrase has the structure [Num+Adj+CL+N], in the analysis window of the classifier, on hearing *ba*, L1ers started to fix more on the mass/plural-expressing pictures instead of the count-noun denoting pictures. On the other hand, when the noun was a mass noun, it is reasonable to observe longer FFD in T1 than in T2 and T3. This is because in T1 the classifier *ba* is a count-biased classifier which led participants to expect count nouns to be the upcoming noun. On reaching the mass noun, participants started to realize that their original analysis was wrong and needed to suppress the dominant count-classifier meaning of *ba* and coerce it to have the less-likely massifier meaning, since mass nouns can only occur with massifiers. This reanalysis procedure caused the longer processing times of mass nouns in T1 than in T2 and T3, since in the mass-biased structures (T2 & T3), *ba* was interpreted as a massifier

which could fit with mass nouns. The observed FFD patterns on count nouns and mass nouns with the classifier *ba* are exactly as I have predicted.

When the classifier was *gen/kuai/pian*, there was no significant interaction between the nominal phrase structures and the mass/count nouns. This finding is reasonable if considering these classifiers are mass-biased classifiers. Since they already have dominant massifier meanings, the mass-biased structures (T2 & T3) or the neutral structure (T1) would have no effect on their processing and interpretations. The only factor affected L1ers' FFD on the nouns is the mass/count status of the nouns: count nouns had longer FFD than mass nouns. This pattern is consistent with the co-occurrence frequency of count/mass nouns and different classifiers. As discussed above and illustrated in Table 23, mass nouns have higher co-occurrence frequency with the classifiers *gen/kuai/pian* than count nouns. Based on the distributional properties of classifier-noun pairs, it is reasonable to observe that count nouns (which have lower co-occurrence frequency) had longer processing times than mass nouns (which have higher co-occurrence frequency) when the classifiers were *gen/kuai/pian*.

For L2ers, the best fit LME model showed that different from L1ers, the nominal phrase structures did not affect their early stage processing of the nouns, only the mass/count status of the nouns, the classifiers and their interactions are included as the fixed factors. When the classifier was *ba*, the difference of FFD between count nouns and mass nouns was not significant. When the classifier was *kuai*, count nouns had marginally significantly longer FFD than mass nouns ( $p = 0.0682$ ). When the classifier was *gen/pian*, FFD on the count nouns was significantly longer than mass nouns. Based on the discussion above we can see that, L2ers' FFD patterns on the nouns were exactly the same as the co-occurrence frequency patterns of the classifier-noun pairs. Count nouns and mass nouns have equal co-occurrence frequency with the classifier *ba*, which lead to the insignificant difference of FFD between count nouns and mass nouns. Mass nouns have higher co-occurrence frequency than count nouns when occurring with the classifiers *gen/kuai/pian*, thus mass nouns had shorter FFD than count nouns when the classifiers were *gen/kuai/pian*. Considering that L2ers can only implicitly learn these classifier-noun relations through daily language experience, it is reasonable to observe that their behaviour patterns were influenced by the distributional properties of the classifier-noun pairs.

In general, the results of the early stage processing of the nouns indicate that the co-occurrence frequency of the classifier-noun pairs had some influences on both L1ers' and L2ers' FFD on the nouns. Shorter FFD was found on the nouns which have higher co-occurrence frequency with the classifier. L1ers were sensitive to the count-bias property of the classifier *ba*, and the difference between the neutral structure (T1) and the mass-biased structures (T2 & T3). However, L2ers failed to exhibit the sensitive to the count-bias property of the classifier *ba*.

### 6.3.2. *The late stage processing of the nouns (total reading times)*

The best fit LME models for both L1ers and L2ers show that the only fixed factor that had an influence on participants' TRT in the ROI of the nouns is the classifier. For both L1ers and L2ers, TRT on the nouns with the classifier *ba* was significantly shorter than with the other three classifiers. This can be explained by the frequency effect. In general, nominal phrases with the classifier *ba* have the higher frequency than nominal phrases with the classifier *gen/kuai/pian* ( $ps < 0.001$ ). This high frequency of nominal phrases with *ba* led to the shorter TRT for both L1ers and L2ers.

To sum up, the late stage processing of the nouns indicate that similar to L1ers, L2ers had no problems of understanding the mass and count nouns in all the three types of nominal phrases. The higher frequency the nominal phrase has, the shorter TRT was needed.

## 6.4. The processing of the phrases

The best fit LME models of TRT in the ROI of the phrases do not include the mass/count status of nouns as the fixed factor for either L1ers or L2ers, which indicates that both L1ers and L2ers had no problems of understanding the nominal phrases no matter with count nouns or mass nouns. This is consistent with the results of the Grammaticality Test that all nominal phrases are grammatical and acceptable to native Mandarin speakers.

The interactions between nominal phrase structures and different classifiers were significant for L1ers, but not for L2ers. For L1ers, the TRT pattern in the ROI of the phrases with *ba* was different from it with the other three classifiers. When the classifier was *ba*, the nominal phrases in T3 had significantly longer TRT than in T1. And the nominal phrases in T2 had marginally significantly longer TRT than in T1. The difference between T2 and T3 was not significant. The finding that TRT on the phrases in T1 was shorter than in T2 and T3 is reasonable if we consider the classifier *ba* as a

count-biased classifier. Putting a count-biased classifier (*ba*) in a mass-biased structure (T2 and T3) would cause some extra time for L1ers to process due to the coercion procedure – participants need to coerce the dominant count-classifier meaning of the count-biased classifier *ba* to a less-likely massifier meaning. Also, TRT in general on the phrases in T3 was longer than in T1 and T2. Recall that nominal phrases in T1 has the structure [Num+CL+Adj+N], in T2 has the structure [Num+Adj+CL+N], and in T3 has the structure [Num+Adj+CL+*de*+N]. TRT is a late stage on-line processing index which can reflect participants' understanding of the linguistic inputs. Since the Grammaticality test results indicate that L1ers had no problems of understanding nominal phrases in all three types of structures, the longer TRT in T3 than in T1 and T2 can only be caused by the extra character in T3 – the modification marker *de*. There are four words in the nominal phrases in T1 and T2, but five words in the nominal phrases in T3. Thus it is reasonable to observe that nominal phrases in T3 had longer processing times than in T1 and T2.

Different from *ba*, the other three classifiers share the same patterns: when the classifier was *gen/kuai/pian*, nominal phrases in T3 had significantly longer TRT than in T1 and T2; the difference between T1 and T2 was not significant. Since the classifiers *gen/kuai/pian* are mass-biased classifiers, different nominal phrase structures would not affect participants' understanding of nominal phrases because there is no conflict between the dominant meanings of the classifiers and the preferred meanings of the structures. It is reasonable to observe no significant difference of TRT on the phrases between T1 and T2. As for the longer TRT in T3, it can be explained by the extra *de* in the nominal phrases in T3.

L1ers' results from the best fit LME model indicate that in general, the properties of different classifiers affected L1ers' processing and understanding of the phrases: putting a count-biased classifier (*ba*) in a mass-biased structure (T2 & T3) caused some processing and understanding difficulties. L1ers were sensitive to the mass/count structures and properties of different classifiers (count/mass-biased properties, preferred count/mass nouns).

Different from L1ers, the best fit LME model for L2ers does not include the interactions between nominal phrase structures and different classifiers. Instead, the best fit LME model for L2ers includes the participants as a random factor, indicating that L2ers had

larger variations than L1ers. The results of the best fit LME model show that, for L2ers, different nominal phrase structures affected TRT in the ROI of the phrases: nominal phrases in T3 had significantly longer TRT than in T1. This can be explained by the extra *de* in nominal phrases in T3. Also, the different classifiers affected on L2ers' late stage processing of the phrases. The phrases with the classifier *ba* had significantly shorter TRT than the phrases with the other three classifiers. This is consistent with the frequency of the nominal phrases with different classifiers. Among the four classifiers, nominal phrases with *ba* have higher frequency than nominal phrases with the other three classifiers ( $ps < 0.01$ ). The differences of the nominal phrase frequency among the other three classifiers *gen/kuai/pian* are not significant. Since L2ers in the current study could only implicitly learn the classifier-noun connections through immersive language input environments, it is reasonable that they were influenced by the distributional properties of nominal phrases. The finding that there is no interaction between nominal phrase structures and different classifiers indicates that the nominal phrase structures affected L2ers' understanding of phrases in the same way with different classifiers. Unlike L1ers, L2ers failed to show sensitive to the count/mass-bias properties of different classifiers. In general, the results of the best fit LME models indicate that different from L1ers, L2ers were more easily affected by the distributional properties of nominal phrases, and less sensitive to the properties of different classifiers.

To sum up, in the current experiment, I manipulated nominal phrase structures (T1 with the structure [Num+CL+Adj+N], T2 with the structure [Num+Adj+CL+N], T3 with the structure [Num+Adj+CL+*de*+N]), different classifiers (count-biased classifiers vs. mass-biased classifiers), and the status of nouns (mass vs. count) to investigate how native Mandarin speakers and high proficiency late Dutch-Mandarin learners react to nominal phrases with different structures, different classifiers and mass/count nouns during on-line processing. The results are close to the predictions: for native Mandarin speakers, the influences from nominal phrase structures and mass/count nouns were only spotted when the classifier was *ba*, but not with the other three classifiers *gen/kuai/pian*. Based on the patterns of processing times, different classifiers can be divided into two groups: *ba* is in one group by itself, while *gen*, *kuai*, and *pian* are in another. This is consistent with different ambiguity types of classifiers: *ba* is a count-biased classifier while *gen/kuai/pian* are mass-biased classifiers. Native Mandarin speakers exhibited

their sensitivity to the count/mass-bias properties of different classifiers, while late Dutch-Mandarin learners failed to do so. This finding indicates that late Dutch-Mandarin learners had not acquired the knowledge about these different ambiguous types of classifiers in Mandarin. Most L2ers' on-line processing behaviours are consistent with the distributional properties of the nominal phrase sequences. These findings illustrate that learning L2 implicitly through daily language input, L2 learners are easily affected by the co-occurrence frequency of the sequences and the frequency of nominal phrases with different classifiers.

Based on the results from the best fit LME models, a brief summary of the FFD and TRT patterns in each ROI for both L1ers and L2ers is illustrated in Table 45.

## Chapter 5 \_ Reading for Comprehension Experiment

Table 45. A brief summary of the results in the Reading for Comprehension experiment

	Native Mandarin speakers (L1ers)		Late Dutch-Mandarin learners (L2ers)	
<b>FFD on classifier</b>	<b>Ba:</b> FFD1<FFD2, FFD1=FFD3, FFD3≤FFD2 (p=0.063)	<b>Ba</b> is a count-biased classifier; sensitive to Adj-CI structure; Parafoveal-on-foveal effect & measurement phrases	<b>Ba:</b> FFD1=FFD3<FFD2	L2ers' specific linguistic input
	<b>Gen:</b> FFD1=FFD2=FFD3	<b>Gen</b> is a mass-biased classifier	<b>Gen:</b> FFD1<FFD2	
	<b>Kuai:</b> FFD1=FFD2=FFD3	<b>Kuai</b> is a mass-biased classifier	<b>Kuai:</b> FFD1=FFD3<FFD2	
	<b>Pian:</b> FFD1=FFD2=FFD3	<b>Pian</b> is a mass-biased classifier	<b>Pian:</b> FFD1=FFD2=FFD3	
<b>TRT on classifier</b>	TRT1=TRT3<TRT2	Frequency effect	TRT1=TRT3<TRT2	Frequency effect
<b>FFD on noun</b>	<b>Ba:</b> Count N: FFD1<FFD2=FFD3 Mass N: FFD1>FFD2=FFD3	Ba is a count-biased classifier which has strong connections with count nouns; sensitive to Adj-CI structure	<b>Ba:</b> FFDcount=FFDmass	Frequency effect; not sensitive to the <b>Ba</b> -N connections
	<b>Gen:</b> FFDcount>FFDmass	Frequency effect	<b>Gen:</b> FFDcount>FFDmass	Frequency effect
	<b>Kuai:</b> FFDcount>FFDmass	Frequency effect	<b>Kuai:</b> FFDcount>FFDmass	Frequency effect
	<b>Pian:</b> FFDcount>FFDmass	Frequency effect	<b>Pian:</b> FFDcount>FFDmass	Frequency effect
	<b>TRT on noun</b>	TRTba<TRTgen=TRTkuai=TRTpian	TRTba<TRTgen=TRTkuai=TRTpian	Frequency effect
<b>TRT on phrase</b>	<b>Ba:</b> TRT1<TRT2=TRT3	<b>Ba</b> is a count-biased classifier & the extra character 'de'	TRT1<TRT3, TRT1=TRT2, TRT2=TRT3	The extra character 'de'
	<b>Gen:</b> TRT1=TRT2<TRT3	The extra character 'de'		
	<b>Kuai:</b> TRT1=TRT2<TRT3	The extra character 'de'	TRTba<TRTgen=TRTkuai=TRTpian	Frequency effect
	<b>Pian:</b> TRT1=TRT2<TRT3	The extra character 'de'		



## VI. General Discussion

I set out to argue that late L2ers can eventually exhibit native-like behaviours during real time processing of unique-to-L2 constructions. Late L2ers do not have fundamentally different representations from native speakers. Their processing difficulties are caused by a deficit in performance but not a deficit in representations.

Two eye-tracking experiments were conducted on high proficiency late Dutch-Mandarin learners and native Mandarin speakers. The results showed that in the Visual World Paradigm experiment (with a comparatively low workload), late Dutch-Mandarin learners exhibited native-like behaviours in making use of the mass/count syntactic cues in predictive processing. However, in the Reading for Comprehension experiment (with a comparatively high workload), late Dutch-Mandarin learners exhibited non-native-like patterns. Late Dutch-Mandarin learners' different behaviours between these two experiments indicated a workload effect, which supports my argument and the Performance Deficit Account (Dekydtspotter, Schwartz and Sprouse, 1996; Hopp, 2006; Schwartz & Sprouse, 1996; White, 2003) which states that late L2ers' on-line processing difficulties are caused by their limited cognitive resources, but not their non-native-like L2 representations. With high L2 proficiency and in a low-workload-demanding task, it is highly possible for late L2ers to display native-like patterns. The native-like behaviours in Dutch-Mandarin learners spotted in the Visual World Paradigm experiment are consistent with the Feature Reassembly Hypothesis (Lardiere, 2009) which states that late L2ers can eventually acquire unique-to-L2 features through the 'feature reassembly' procedure. However, this finding is inconsistent with the Morphological Congruency Hypothesis (Jiang et al., 2011) which states that late L2ers cannot have native-like representations of unique-to-L2 constructions due to the negative L1-transfer.

In this chapter, an overview of this current research is presented first, followed by the detailed discussions of the main findings.

## 1. Overview of the current research

In order to look into whether late L2ers can acquire unique-to-L2 knowledge through implicit learning, and make use of this knowledge to build anticipations during real time processing in a native-like manner, the current research conducted a Visual World Paradigm experiment and a Reading for Comprehension experiment on high proficiency late Dutch-Mandarin learners and native Mandarin speakers. According to Cheng & Sybesma (1998, 1999), different from languages like English and Dutch in which the mass/count distinction is encoded in plural morphemes and articles, in Mandarin, the mass/count distinction is marked in the classifier system. They distinguish massifiers from count-classifiers and show that nominal phrases with massifiers have specific structures: the modification marker ‘*de*’ can only be inserted after a massifier but not a count-classifier, and certain adjectives (e.g. big and small) can be added before or after a massifier, but can only occur after a count-classifier. In addition, this difference between count-classifiers and massifiers, and the corresponding mass/count syntactic cues are implicit to both native Mandarin speakers and Dutch-Mandarin learners. Native Mandarin speakers use these constructions automatically and unconsciously in daily life, without being able to realize it and verbalize specific rules. As for late Dutch-Mandarin learners, they can only explicitly learn the obligatory appearance of a classifier between a numeral and a noun, and some high-frequency classifier-noun pairings, which are taught in Mandarin-teaching classes and textbooks. No explicit and specific instruction about the differences between count-classifiers and massifiers, and the mass/count syntactic cues can be found in Mandarin-teaching textbooks or classes. Dutch-Mandarin learners can only acquire these unique-to-L2 constructions (i.e. the difference between count-classifiers and massifiers, and the mass/count syntactic cues) implicitly from an immersive Mandarin-dominant environment. Exploring late Dutch-Mandarin learners’ on-line use of classifiers and the mass/count syntactic cues helped us gain some understandings of late L2ers’ on-line processing of unique-to-L2 constructions which can only be acquired through implicit learning.

The Visual World Paradigm experiment manipulated the Cl-Adj word order and found that even though Dutch-Mandarin learners were slower than native Mandarin speakers, both groups can use the mass/count syntactic cue (i.e. the Adj-Cl word order) on-line to make mass/plural-preferred anticipations during real time processing. Nominal phrases

with the Adj-Cl order triggered more fixations on the mass/plural-expressing pictures than the count-expressing pictures on both native Mandarin speakers and Dutch-Mandarin learners before the onset of the nouns. In addition, the Visual World Paradigm experiment also found that different classifiers evoked different fixation patterns for native Mandarin speakers: the onset of the classifier *ba* triggered more fixations in the count-noun denoting pictures, while on hearing *gen/kuai*, no preference was spotted. This classifier division effect during on-line processing was spotted only on native Mandarin speakers but not on Dutch-Mandarin learners. Moreover, different from native Mandarin speakers who used the Adj-Cl structure to build mass/plural-preferred anticipations and fixed on mass/plural-expressing pictures immediately after the onset of the Adj, Dutch-Mandarin learners were attracted by the lexical meanings of the Adj. They directed their attentions to the pictures containing Adj-consistent-size objects instead of mass/plural-expressing pictures on hearing the Adj. This finding indicated that for the same Adj, native Mandarin speakers were more sensitive to its syntactic position and function while Dutch-Mandarin learners were more sensitive to its lexical-semantic meaning.

To further look into the possible interactions between the mass/count syntactic cues and different classifiers, and how Dutch-Mandarin learners react in a comparatively higher workload required paradigm, the Reading for Comprehension experiment was conducted, in which both the Adj-Cl word order and the insertion of *de* were manipulated. The results showed that, apart from the mass/count syntactic cues, the different ambiguity types of classifiers also affected native Mandarin speakers' on-line processing: when there was a mismatch between the preferred meaning of the nominal phrase structure and the dominant meaning of the classifier, longer processing times were found on the classifier. Also, when the noun following the classifier was not a preferred associated noun, longer processing times were observed on the noun. Dutch-Mandarin learners exhibited native-like patterns at the late stage processing of nominal phrases (illustrated through the total reading times), but non-native-like patterns at the early stage (illustrated through the first fixation durations). Compared to native Mandarin speakers who were affected by both the mass/count syntactic cues and the count/mass bias properties of classifiers, Dutch-Mandarin learners were strongly affected by the co-occurrence frequency of the sequences and the frequency of nominal phrases with different classifiers.

Taken together, the main findings of the two eye-tracking experiments are summarized as follows.

- Dutch-Mandarin learners were slower than native Mandarin speakers in making use of the Adj-Cl word order to build mass/plural-preferred interpretations.
- Dutch-Mandarin learners were sensitive to the Adj-Cl order; they can use this implicit unique-to-L2 knowledge to make mass/plural-preferred anticipations during real time processing.
- Dutch-Mandarin learners were not sensitive to the mass/count-bias properties of classifiers.
- Dutch-Mandarin learners were more sensitive to the lexical meanings of Adj than the interpretive properties of its syntactic position.

In the following part of this chapter, detailed discussions about these four main findings are presented in Section 2 to Section 5, with respect to previous studies and relevant theories. The conclusion and contributions of the current research are presented in Section 6.

## **2. L2ers' slowness compared to L1ers'**

The Visual World Paradigm experiment revealed that Dutch-Mandarin learners were slower than native Mandarin speakers in making use of the Adj-Cl order to build mass/plural-preferred anticipations: on hearing the adjective directly after the numeral, native Mandarin speakers immediately predicted the upcoming item to be a massifier and landed their fixations to the mass/plural-expressing pictures, while Dutch-Mandarin learners started to direct their attention to these pictures only after the onset of the classifier. The Reading for Comprehension experiment revealed that in general, Dutch-Mandarin learners were slower than native Mandarin speakers in language processing: Dutch-Mandarin learners' first fixation durations and total reading times on each item of nominal phrases were longer than native Mandarin speakers'. However, in this current research, I argue that the slowness of Dutch-Mandarin learners in making use of the Adj-Cl order to build anticipations in the Visual World Paradigm experiment is not the

consequence of the general processing slowness of Dutch-Mandarin learners found in the Reading for Comprehension experiment and a number of previous studies (e.g. Hahne & Friederici, 2001; Lew-Williams & Fernald, 2010; Sanders & Neville, 2003). In the Visual World Paradigm experiment, even though in Condition 3 (in which the nominal phrase has the structure [Num+Adj+Cl+N]) Dutch-Mandarin learners exhibited native-like fixation patterns on the Adj-Cl word order in the analysis window of the classifier, which was slower than native Mandarin speakers who exhibited the same pattern in the analysis window of the adjective, they also exhibited this Adj-sensitive pattern in the analysis window of the adjective in Condition 2 ([Num+Cl+Adj+N]). The consistency of Dutch-Mandarin learners' sensitivity to the lexical information of the adjectives in Condition 2 and Condition 3 indicated that the late appearance of the native-like fixation patterns on hearing the Adj-Cl word order in Condition 3 is not due to Dutch-Mandarin learners' slow reaction in general language input processing, rather it is the result of their different predictive processing patterns from native Mandarin speakers: compared to native Mandarin speakers, they need more time/information to build anticipations. On hearing the adjective directly following a numeral, native Mandarin speakers already started to build a mass/plural-preferred nominal phrase structure and predicted the upcoming item to be a massifier. As a consequence, they started to land their fixations on the mass-/plural-expressing pictures at this point. Dutch-Mandarin learners, on the other hand, have not yet realized that the nominal phrase will have a mass-preferred structure at the onset of the adjective. On hearing the adjective directly following a numeral, instead of building a mass-preferred structure like native Mandarin speakers, they integrated the lexical information of the adjective and consequently directed their fixations to the pictures containing Adj-consistent-size objects. Thus compared to native Mandarin speakers who can take advantage of the direct appearance of an adjective after a numeral to build a mass-preferred anticipation, Dutch-Mandarin learners need the whole structure [Num+Adj+Cl] to make a mass-preferred prediction.

The longer processing times (first fixation durations and total reading times) for Dutch-Mandarin learners compared to native Mandarin speakers in the Reading for Comprehension experiment in the current study have been corroborated by several previous studies (e.g. Hahne & Friederici, 2001; Lew-Williams & Fernald, 2010; Sanders & Neville, 2003). As mentioned in Chapter 1, studies in SLA have found that

L2ers are generally slower than L1ers in both on-line and off-line sentence processing with a variety of tasks and materials (Trenkic, Mirkovic, & Altmann, 2013). This general ‘slowness’ phenomenon of L2ers is predicted by both the Fundamental Difference Approach and the Fundamental Similarity Approach as have discussed in Chapter 1. In the current research I argue that even though both the Fundamental Difference Approach and the Fundamental Similarity Approach can explain the general slow processing by late L2ers which was observed in the Reading for Comprehension experiment, my results support the Fundamental Similarity Approach but not the Fundamental Difference Approach if we consider the results from both of two eye-tracking experiments.

Combining the findings from the two eye-tracking experiments, it is interesting to find that the general slower processing of Dutch-Mandarin learners observed in the Reading for Comprehension experiment was not found in the Visual World Paradigm experiment. Dutch-Mandarin learners exhibited longer first fixation durations and total reading times on each item of nominal phrase than native Mandarin speakers in the Reading for Comprehension experiment. In the Visual World Paradigm experiment, however, the onset of informative cues evoked both native Mandarin speakers and Dutch-Mandarin learners' change of fixation distributions around the same time. In general, there are three informative cues in the Visual World Paradigm experiment: the count-biased classifier *ba* which has strong connections with associated count nouns, the Adj-Cl word order which triggers mass/plural-preferred interpretations, and the nouns which finally determine the target picture. The results of the Visual World Paradigm experiment showed that native Mandarin speakers were sensitive to all three of these cues: they focused on count-noun denoting pictures on hearing the classifier *ba*, landed more fixations on the mass/plural-expressing pictures than the individual count-expressing pictures on hearing the Adj-Cl order, and fixed on the target picture on the onset of the corresponding noun. Dutch-Mandarin learners, on the other hand, exhibited different patterns to these three cues: the onset of the classifier *ba* did not evoke their attention to the count-noun denoting pictures, indicating that they were not sensitive to the count-bias property of the classifier *ba*, and its close connections with associated count nouns; the onset of the adjective always triggered their fixations to the Adj-consistent-size picture regardless of the position of the Adj (the Cl-Adj order in Condition 2 vs. the Adj-Cl order in Condition 3), indicating that they were more sensitive to the lexical information of adjectives than the mass-biased structure with the Adj-Cl order; the onset

of the nouns directed their fixations to the target picture. Among the three informative cues, only the onset of nouns evoked the same patterns on native Mandarin speakers and Dutch-Mandarin learners. And on hearing nouns, both of them shifted their fixations to the target picture around the same time – approximately 200ms after the onset of the nouns. Thus there is no evidence of general processing slow reactions of Dutch-Mandarin learners compared to native Mandarin speakers in the Visual World Paradigm experiment. The slowness in assigning the mass/plural-preferred interpretations to the Adj-CI word order observed in the Visual World Paradigm experiment is different from the slowness in general language input processing observed in the Reading for Comprehension experiment.

Reviewing previous Visual World Paradigm studies which investigated L2ers' predictive processing of unique-to-L2 constructions, I found that no general processing slow reactions have been observed in highly advanced L2ers when compared to L1ers (Hopp, 2012, 2015; Lau & Grüter, 2015). The general slower processing behaviours in L2ers than L1ers were only found in Trenkic et al. (2013) in which L2ers were at an intermediate L2 level. Considering the fact that the late Dutch-Mandarin learners who participated in the current study were high proficiency L2ers, it is unsurprising that they exhibited no general slow processing reactions compared to native Mandarin speakers in the Visual World Paradigm experiment.

However, the question does arise: why did the same group of high proficiency late Dutch-Mandarin learners exhibit general processing slow behaviours in the Reading for Comprehension experiment but not in the Visual World Paradigm experiment? One possible explanation is the difference between the two paradigms. As discussed in Chapter 3, the two experimental paradigms differ with each other in three ways: the material input modality, the task request, and the salience of the materials. In general, the Visual World Paradigm experiment requires a lower workload on participants than the Reading for Comprehension experiment, and consequently, it is reasonable to expect general slow processing behaviours on L2ers in the Reading for Comprehension task (which has a comparatively high workload) but not in the Visual World Paradigm (which has a comparatively low workload).

This contrast between the results of the Visual World Paradigm experiment and the Reading for Comprehension experiment supports the Performance Deficit Account but

not the Representation Deficit Account. Recall that the Representation Deficit Account states that L2ers' slow reactions are caused by a deficit in L2 representations. Thus no matter whether the paradigm demands a high or low workload on L2ers, they should always exhibit slower processing behaviours compared to L1ers due to their deficient L2 representations. Thus the Representation Deficit Account predicts that high proficiency late Dutch-Mandarin learners will exhibit slower behaviours than native Mandarin speakers in the Visual World Paradigm experiment. And further, their slowness compared to native Mandarin speakers should be more significant in the Reading for Comprehension experiment than in the Visual World Paradigm experiment since the former one requires a comparatively higher workload. On the other hand, the Performance Deficit Account claims that it is L2ers' limited cognitive resources that lead to their slow reactions but not their deficient representations. Once the task requires a low workload, high proficiency L2ers can exhibit native-like behaviours since they already acquired native-like representations. Thus the Performance Deficit Account predicts that it is more likely to observe high proficiency late Dutch-Mandarin learners' slower reactions compared to native Mandarin speakers in the Reading for Comprehension experiment than in the Visual World Paradigm experiment due to the higher workload the former one required.

The finding that high proficiency late Dutch-Mandarin learners exhibited slower processing behaviours compared to native Mandarin speakers only in the Reading for Comprehension experiment but not in the Visual World Paradigm experiment is consistent with the prediction based on the Performance Deficit Account. Thus the results in the current research support the Performance Deficit Account, but provide evidence against the Representation Deficit Account.

### **3. L2ers' sensitivity to the Adj-Cl word order**

Even though late Dutch-Mandarin learners were easily attracted by the lexical information of adjectives (they always directed their fixations to the Adj-consistent pictures on hearing the Adj in the Visual World Paradigm experiment), they exhibited sensitivity to the mass-biased structure (the Adj-Cl word order) in the Visual World Paradigm experiment: similar to native Mandarin speakers, on hearing the Adj-Cl order,



Dutch-Mandarin learners fixed more on the mass/plural-expressing pictures than the individual count-expressing picture. This native-like pattern indicates that high proficiency late Dutch-Mandarin learners have acquired the mass/count syntactic cues through implicit learning: the Adj-CI word order triggered their anticipations for a mass/plural-preferred interpretation. This finding is consistent with several previous studies which have shown that late L2ers exhibit native-like behaviours during on-line processing (e.g. Hopp, 2006; Jackson & Dussias, 2009; Yao & Chen, 2016). Also, as mentioned in Chapter 1, previous Visual World Paradigm studies which investigated late L2ers' predictive processing of unique-to-L2 knowledge found that late L2ers can use unique-to-L2 constructions to build anticipations during real time processing (e.g. Hopp, 2012, 2015; Trenkic et al., 2013). Thus it is reasonable to observe native-like behaviours on high proficiency late Dutch-Mandarin learners in the Visual World Paradigm experiment in the current research. Further, this finding offers supportive evidence to the Feature Reassembly Hypothesis (Lardiere, 2009), but against the Morphological Congruency Hypothesis (Jiang et al., 2011). Recall that, as discussed in Chapter 1, the Morphological Congruency Hypothesis (Jiang et al., 2011) claims that it is very difficult, if possible at all, for late L2ers to acquire unique-to-L2 constructions. On the other hand, the Feature Reassembly Hypothesis (Lardiere, 2009) proposed that unique-to-L2 knowledge can be acquired by late L2ers through feature reassembly. The results of the Visual World Paradigm experiment in the current research showed that high proficiency late Dutch-Mandarin learners can use the mass/count syntactic cue (i.e. the Adj-CI word order) in predictive processing, and that they exhibited native-like behaviours when hearing the Adj-CI structure. As discussed in Chapter 2, the classifier system and the mass/count syntactic cues are unique-to-L2 constructions for Dutch-Mandarin learners since they only exist in Mandarin but not in Dutch. Thus the findings in the Visual World Paradigm experiment indicate that late L2ers can acquire unique-to-L2 constructions and exhibit native-like behaviours in real time processing. This result is consistent with the Feature Reassembly Hypothesis (Lardiere, 2009).

However, native-like behaviours were not found in Dutch-Mandarin learners at the early stage of on-line processing in the Reading for Comprehension experiment. At the early stage processing of the classifiers, treating *ba* as a count-biased classifier, native Mandarin speakers exhibited their sensitivity to the mass/count structures: the first fixation durations on *ba* in T2 (in which the nominal phrase has the structure

[Num+Adj+Cl+N]) were significantly longer than in T1 ([Num+Cl+Adj+N]). The longer fixation on *ba* in T2 than in T1 was caused by the conflict between the mass-preferred interpretation from the structure and the count-biased dominant meaning from the classifier *ba*. Different from native Mandarin speakers, Dutch-Mandarin learners did not exhibit sensitivity to the mass-biased structures and the count-bias property of the classifier *ba*. Most of their first fixation duration patterns can be explained by the co-occurrence frequency of the sequences in different types of structures and the distribution properties of the classifier-noun pairings: the higher frequency of the sequences in T1 and T3 ([Num+Adj+Cl+de+N]) compared to the sequences in T2 led to the shorter first fixation durations on the classifiers in T1 and T3 than in T2. At the early stage of processing of the nouns, native Mandarin speakers exhibited a distinction between count nouns and mass nouns when the classifier was *ba*: the first fixation durations on count nouns in T1 were significantly shorter than T2 and T3, while the first fixation durations on mass nouns in T1 were significantly longer than T2 and T3. This is because of the strong connection between the count-classifier *ba* and its associated count nouns. The neutral structure in T1 led the count-biased classifier *ba* to cause a strong anticipation for the upcoming noun to be a count noun, which consequently facilitated the processing of count nouns while suppressing the activation of mass nouns. These results indicate that native Mandarin speakers were sensitive to the Adj-Cl word order as well as the count-bias property of the classifier *ba*. As for Dutch-Mandarin learners, the different nominal phrase structures did not affect their first fixation durations on count/mass nouns when the classifier was *ba*. In all three types of nominal phrase structures, their first fixation durations on count nouns showed no significant difference with mass nouns, which is consistent with the co-occurrence frequency of count and mass nouns and the classifier *ba* (as discussed in Chapter 5, count nouns and mass nouns have a similar co-occurrence frequency occurring with the classifier *ba*). In general, the results of the Reading for Comprehension experiment showed that late Dutch-Mandarin learners exhibited non-native-like behaviours during on-line processing: they were not sensitive to the count-bias property of the classifier *ba*, or its strong connections with associated count nouns. Instead, Dutch-Mandarin learners were strongly influenced by the co-occurrence frequency of the sequences.

The non-native-like behaviours on late L2ers have been found by a number of previous studies (Chen et al., 2007; Ionin et al., 2012; Jiang, 2004, 2007; Kotz, 2009). This result

can be explained by both the Representation Deficit Account and the Performance Deficit Account. The former states that after puberty it is very difficult for L2ers to acquire native-like L2 representations. And as a consequence, late L2ers cannot exhibit native-like behaviours due to their deficient representations. On the other hand, the latter argues that the non-native-like behaviours on late L2ers are the results of their limited cognitive resources and L2 proficiency. With increased L2 proficiency and a low workload, late L2ers can eventually exhibit native-like performance. Even the results in the Reading for Comprehension experiment can be explained by both the Representation Deficit Account and the Performance Deficit Account, in this current research I argue that, the results of the two eye-tracking experiments generally support the Performance Deficit Account that late L2ers can eventually exhibit native-like behaviours, they do not have deficient representations.

The comparison of the results in the two eye-tracking experiments illustrated that Dutch-Mandarin learners' native-like behaviour patterns were only spotted in the Visual World Paradigm experiment but not in the Reading for Comprehension experiment. There are two possibilities to explain this inconsistency between the results of the two experiments in the current research. First of all, the Visual World Paradigm experiment looked into participants' on-line use of the mass/count syntactic cue (i.e. the Adj-Cl word order) while the Reading for Comprehension experiment investigated the influence of the mass/count syntactic cues as well as the properties of different types of ambiguous classifiers. The Visual World Paradigm experiment revealed that Dutch-Mandarin learners were not sensitive to the count-bias properties of the classifier *ba* and its strong connection with associated count nouns, which has also been found in the Reading for Comprehension experiment. Thus Dutch-Mandarin learners' non-native-like behaviours in the Reading for Comprehension experiment may be caused by their non-native mastery of the mass/count-bias properties of classifiers, but not their use of mass/count syntactic cues. In other words, late Dutch-Mandarin learners may have already acquired the mass/count syntactic cues but not the count/mass-bias properties of classifiers. The second possible explanation for the inconsistency is the workload effect. As discussed above, the Visual World Paradigm experiment and the Reading for Comprehension experiment differ in their workloads: compared to the Visual World Paradigm, the Reading for Comprehension experiment requires a higher workload, which may lead to the non-native-like behaviours on Dutch-Mandarin learners. If this is on the right track,

then the results from the two experiments in the current research offer some evidence to support the Performance Deficit Account. As discussed in Chapter 1, the Performance Deficit Account (Dekydtspotter, Schwartz and Sprouse, 1996; Hopp, 2006; White, 2003) predicts that late L2ers' on-line processing is limited by their performance but not their representations. Under high-workload-demanded conditions, and with limited cognitive processing resources, late L2ers may exhibit non-native-like behaviours even if they have already acquired the L2 knowledge and had native-like representations. When with high L2 proficiency and a low workload, it is highly possible for late L2ers to exhibit native-like behaviours in on-line processing. The workload effect observed in the current research is consistent with this prediction: high proficiency Dutch-Mandarin learners exhibited native-like behaviours in the low-workload-demanding experiment (the Visual World Paradigm experiment) but non-native-like behaviours in the high-workload-demanding experiment (the Reading for Comprehension experiment). Thus, combining the findings of the two experiments I argue that the results of this current research support my argument and the Performance Deficit Account that late L2ers can acquire unique-to-L2 constructions and make use of them during on-line processing. Their processing difficulties are caused by limited cognitive resources but not deficient representations.

#### 4. L2ers' problems with different classifiers

As discussed before, in this current research, Dutch-Mandarin learners can process mass/count syntactic cues on-line (revealed in the Visual World Paradigm experiment), but they exhibited some problems with the count/mass-bias properties of classifiers (revealed in the Reading for Comprehension experiment). It should be noted that both the mass/count syntactic cues and the count/mass-bias properties of classifiers are implicit and unique-to-L2 constructions to Dutch-Mandarin learners. But why did they only exhibit native-like behaviours on the mass/count syntactic cues but not the mass/count-bias properties of classifiers? One possibility to explain this finding is the order of L2 acquisition.

Previous studies which focus on Chinese children's acquisition of classifiers found that children under 5-years-old first acquire the general classifier *ge* and use it as a 'syntactic place-holder' (Fang, 1985; Hu, 1993a). They often overuse the general classifier *ge* with

any noun and ignore the restrictions of the classifier-noun pairings, which is unacceptable and inappropriate in adult language. They are familiar with the obligatory appearance of a classifier between a numeral and a noun, but not with the semantic relations between the properties of the classifiers and the nouns. According to Fang (1985) and Ying et al. (1983), it is only around the age of five that children start to realize there is a semantic co-occurrence restriction between a classifier and a noun (i.e. a classifier selects a group of nouns which share a certain size/shape/category, etc.). Thus in L1 children's acquisition of classifiers, they acquire the syntactic structure of classifiers first, and then the semantic relations between classifiers and nouns. This acquisition order of classifiers has also been observed in L2ers. Previous studies which looked into the acquisition of classifiers by L2ers found that although they are aware of the obligatory presence of a classifier in a number phrase, L2ers often fail to select the appropriate classifier (Gao, 2009). They often use the general classifier *ge* to fill the syntactic position of classifier, or randomly choose a non-matching classifier (Liang, 2008; Polio, 1994). For L2ers, the syntactic structure of classifiers is more easily acquired than the semantic restrictions of classifier-noun pairings (Gong, 2010).

Based on these previous studies on L1 children's and L2ers' acquisition of classifiers we can say that classifier syntax is more easily acquired than the restrictions of classifier-noun pairings, and thus it is reasonable to expect that late Dutch-Mandarin learners exhibit native-like behaviours on the mass/count syntactic cues but not the classifier-noun restrictions in the current research. Also, as discussed in Chapter 2, the count/mass syntactic cues and count/mass-bias properties of classifiers are implicit and unique-to-L2 constructions to Dutch-Mandarin learners, which can only be acquired through implicit learning. According to Krashen (2005), compared to explicit learning, implicit learning is a product of a subconscious process which is very similar to the process children undergo when they acquire their L1. Thus in the current study, the observation of L1-children-like patterns in the use of classifiers on late Dutch-Mandarin learners is unsurprising, given that they can only acquire the knowledge of the mass/count classifier system through implicit learning.

Furthermore, the findings of the current study implied that late L2ers acquire the classifiers in the same pattern as L1 children do, even though there is no classifier instantiated in their native languages. This is inconsistent with the claim of the Critical Period Hypothesis. The Critical Period Hypothesis introduced by Lenneberg (1967)

states that due to maturational reasons, language acquisition is constrained by language learners' ages. If language acquisition does not occur by puberty, some aspects of language (e.g. grammar, morphosyntax, etc.) cannot be fully mastered by language learners. Extended to second language acquisition, the Critical Period Hypothesis holds that after puberty, L2s can no longer be acquired by the innate language acquisition device and as implicit and proceduralized knowledge, but must be learned incompletely through explicit instructions and long-time practice (e.g. Krashen, 1994; MacWhinney, 2005). Thus, the Critical Period Hypothesis predicts that due to maturational reasons, late L2ers cannot acquire L2 knowledge through implicit learning. In the current study, however, the findings that late Dutch-Mandarin learners have acquired the mass/count syntactic cues through implicit learning, and can use this knowledge in predictive processing, and that they go through the same stages as L1 children do in classifier acquisition, indicate that they can implicitly learn L2 knowledge in the same way L1 children do even after puberty. This is consistent with previous studies that L1 and L2 language development are similar; all language learners go through similar systematic stages in their acquisition and are guided by the same underlying principles (e.g. Flynn, 1996; Gong, 2010; White, 1989). Thus the results in the current study argue against the Critical Period Hypothesis.

The L2ers who participated in the current research were high proficiency late Dutch-Mandarin learners, but not near-native or end-state L2ers. Their problems with the count/mass-bias properties of classifiers may disappear after a long time of immersion in a Mandarin-dominant environment. Further research needs to be done to explore how late L2ers with lengthy immersion in an L2 environment would use the count/mass-bias properties of classifiers in real time processing.

## 5. L2ers' on-line use of Adj information

Another interesting finding in the current research is the different reactions to the adjectives between native Mandarin speakers and Dutch-Mandarin learners. In the Visual World Paradigm experiment, when nominal phrases have the structure [Num+Cl+Adj+N] (Condition 2), the onset of the Adj did not affect native Mandarin speakers' fixation distributions, while it triggered Dutch-Mandarin learners' attention to

the pictures which contained objects organized in Adj-consistent sizes. When the nominal phrases have the structure [Num+Adj+Cl+N] (Condition 3), the appearance of the Adj directly after the numeral evoked native Mandarin speakers' anticipations of mass/plural-preferred interpretations. They landed fixations to the mass/plural-expressing picture on hearing the Adj. Dutch-Mandarin learners, on the other hand, fixed on the Adj-consistent-size picture on hearing the Adj. The different positions of an adjective (Cl-Adj vs. Adj-Cl) affected native Mandarin speakers' predictive processing but not Dutch-Mandarin learners'. Compared to native Mandarin speakers who were sensitive to the position and the corresponding structures of an adjective, Dutch-Mandarin learners were more sensitive to the lexical information encoded in the adjective. This is not a surprising finding if we look through previous studies: late L2ers' sensitivity to lexical information rather than syntactic information has been found in several previous studies (Felser et al., 2003; Lau & Gruter, 2015; Lew-Williams & Fernald, 2009; Papadopoulou & Clahsen, 2003; Roberts & Felser, 2011).

This finding can be explained by the Representation Deficit Account which states that late L2ers have deficient L2 representations. During on-line sentence processing, they need to rely on surface information to compensate for their deficient representations. Thus in real time processing, late L2ers may exhibit more sensitivity to lexical/semantic information than syntactic information, since the lexical/semantic information is shallower and easier to be acquired compared to the syntactic information.

It could be possible that L2ers' over-reliance on surface information decreases as their L2 proficiency increases, since more native-like representations will be involved in on-line processing when L2ers reach a higher proficiency level. Further research needs to be done to test this possibility.

## 6. Conclusions and Contributions

The results of the two eye-tracking experiments showed that high proficiency late Dutch-Mandarin learners can acquire unique-to-L2 constructions (i.e. the mass/count syntactic cues) through implicit learning, and like native Mandarin speakers, automatically use them to build anticipations during real time processing. Also, tasks with different

workloads affect late Dutch-Mandarin learners' performance differently. They exhibited native-like patterns in the Visual World Paradigm experiment (the low-workload task), but non-native-like patterns in the Reading for Comprehension experiment (the high-workload task). These findings firstly support my argument and the Performance Deficit Account that there is no fundamental difference of representations between L1ers and L2ers. Late L2ers can ultimately acquire L2 representations in a native-like manner no matter whether the construction is instantiated in L1 or not. The observed L2 processing difficulties are caused by late L2ers' performance (i.e. limited cognitive resources) but not their deficient representations. In a low-workload-demanding task, it is very likely for high proficiency L2ers to display native-like behaviours. Furthermore, these findings also support the Feature Reassembly Hypothesis which states that L2ers can eventually acquire unique-to-L2 knowledge even after puberty.

However, some of the results in the current research can only be explained by the Representation Deficit Account. In the Visual World Paradigm experiment, compared to native Mandarin speakers, Dutch-Mandarin learners were more sensitive to the lexical information of an adjective than its syntactic information, which is predicted by the Representation Deficit Account. Considering the fact that the late Dutch-Mandarin learners in the current research were high proficiency L2ers, they may still in the process of Mandarin-acquisition, and have some 'deficient' representations which may turn into 'fully native-like' representations with increased L2 proficiency. Further research needs to be done on near-native or end-state Dutch-Mandarin learners to test their mastery of adjectives.

To sum up, the main findings in the current study and their theoretical contributions are summarized in Table 46.

*Table 46. A summary of main findings*

<b>Main findings</b>	<b>Support/Against</b>	<b>Theories</b>
L2ers' general slower processing than L1ers' in the Reading for comprehension experiment but not in the Visual World Paradigm experiment	support	Performance Deficit Account
	Against	Representation Deficit Account



L2ers' sensitivity to the Adj-CI word order	support	Feature Reassembly Hypothesis
	Against	Morphological Congruency Hypothesis
L2ers' similar acquisition pattern with L1 children's: they were sensitive to the mass/count syntax but not to the mass/count-bias of classifiers and restrictions of CI-N pairings	Against	Critical Period Hypothesis
L2ers were more sensitive to the lexical information of an adjective than its syntactic information	Support	Representation Decifit Account

---

In my opinion, L2 acquisition and processing is a long and complicated continuum. With higher proficiency levels and longer practice, more native-like representations and behaviours are expected on late L2ers. Even though part of the results in the current research can be explained by the Representation Deficit Account, based on the general results of the two eye-tracking experiments, I argue that there is no fundamental difference between L1ers and L2ers. Late L2ers can ultimately acquire full L2 representations and exhibit native-like performance.

This current research offers some experimental evidence to support the claim that there is implicit learning in SLA. As discussed in Chapter 1, previous studies which investigated implicit L2 learning mainly used artificial or semi-artificial materials and some off-line paradigms. Very few studies used real language materials and on-line processing paradigms to explore L2ers' implicit learning of unique-to-L2 knowledge. The results of the current study indicate that implicit learning could happen to late L2ers. They can implicitly acquire some unique-to-L2 constructions in real natural languages through an immersive experience in a L2-dominant environment.

In addition, this current research offers some insights about native Mandarin speakers and L2-Mandarin learners' on-line processing of mass/count structures. As discussed in

Chapter 3, previous studies which investigated L1ers' and L2ers' processing of classifiers in Mandarin mostly focused on participants' on-line and off-line reactions to different classifiers. So far, Li, Barner, & Huang (2008) is the only study which used off-line tests to explore native Mandarin speakers' understanding of different classifiers and different nominal phrase structures (the mass/count structures). Until now, it was unclear how native Mandarin speakers and L2-Mandarin learners process nominal phrases with different classifiers and different structures in real-time. The current study filled this gap and found that different structures as well as different classifiers affected both L1ers' and L2ers' on-line processing. The Adj-CI word order was used by both groups as a mass/plural-preferred syntactic cue in the Visual World Paradigm experiment, and the insertion of the modification marker *de* after the classifier always leads to a measurement interpretation in the Reading for Comprehension experiment. Classifiers' count/mass-bias properties have an interaction with different nominal phrase structures: when there is a conflict between the biased meaning evoked by the nominal phrase structure and the dominant meaning of an ambiguous classifier, L1ers exhibited longer processing times on the classifier. L2ers were not sensitive to the mass/count-bias properties of classifiers and their connections with associated nouns.

The current research also makes some contributions to L2 education practice. Even though the findings indicate that cross-language (dis)similarity does not affect late L2ers' acquisition and that they can eventually acquire unique-to-L2 knowledge and exhibit native-like behaviours during on-line processing, this is only achieved after an extended period of implicit learning. Even though the L2 participants in the current research were high proficiency Dutch-Mandarin learners, they were still not sensitive to the count/mass-bias properties of classifiers. According to Tolentino & Tokowicz (2014), an explicit and rule-involved instruction is more efficient than implicit and random exposure for language learning. Thus some explicit and specific instructions on these constructions (e.g. the mass/count-bias properties of classifiers) may be helpful for L2-Mandarin learners to quickly acquire and automatically use this knowledge during on-line processing. L2 education practice could be more effective based on the findings in the current research.

## Appendix A. Materials

### A.1. Typical Mass nouns and Count nouns

Mass nouns in Mandarin	English translation
石子	Pebble
绳子	String
头发	Hair
木棍	Stick
树枝	Branch
香肠	sausage
面包	bread
玻璃	glass
肥皂	Soap
橡皮	Eraser
蛋糕	Cake
奶糖	Candy
黄油	Butter
奶酪	Cheese
Count nouns in Mandarin	English translation
扇子	Fan
梳子	Comb
伞	Umbrella
叉子	Fork
勺子	Spoon
刀	Knife
钥匙	Key
铜锁	Locker
尺子	Ruler
铅笔	Pencil
雪糕	Popsicle
萝卜	Radish
黄瓜	Cucumber
骨头	Bone
香蕉	Banana
油条	Fritters in a long shape
麻花	Fried twist
树叶	Leaf
光盘	CD
花瓣	Petal
羽毛	Feather

---

尿布	Diaper
面膜	Mask
画板	Sketchpad
手绢	Handkerchief
方砖	Squared brick
积木	Building block
玉佩	Jade

---

## A.2. Sentences in the Reading for Comprehension Experiment

### Condition 1: Num+CL+Adj (small/big)+Noun

#### sentences

盒子里装着\*一\*把\*小(大)\*扇子\*, 那盒子是平遥漆盒。  
镜子前摆着\*一\*把\*小(大)\*梳子\*, 那镜子是两面镜。  
衣架上挂着\*一\*把\*小(大)\*伞\*, 那衣架是从德国进口的。  
桌子上摆着\*一\*把\*小(大)\*叉子\*, 那桌子是大理石做的。  
碗里放着\*一\*把\*小(大)\*勺子\*, 那碗是专门给宝宝吃饭的。  
地板上掉着\*一\*把\*小(大)\*刀\*, 那地板是新铺的实木地板。  
墙上挂着\*一\*把\*小(大)\*钥匙\*, 那面墙只在下午能晒到太阳。  
抽屉里放着\*一\*把\*小(大)\*铜锁\*, 那个抽屉很老几乎要坏掉了。  
书包里装着\*一\*把\*小(大)\*尺子\*, 那书包是奶奶给弟弟买的。  
笔袋里装着\*一\*根\*小(大)\*铅笔\*, 那笔袋是大卫送给丽丽的。  
冰箱里放着\*一\*根\*小(大)\*雪糕\*, 那冰箱是丽丽从网上买的。  
冰箱里摆着\*一\*根\*小(大)\*萝卜\*, 那冰箱是大卫新买的。  
盘子里放着\*一\*根\*小(大)\*黄瓜\*, 那盘子是妈妈用来装瓜果的。  
盘子里放着\*一\*根\*小(大)\*骨头\*, 那盘子是用来给狗狗喂食的。  
袋子里装着\*一\*根\*小(大)\*香蕉\*, 那袋子是绿色环保的购物袋。  
碗里放着\*一\*根\*小(大)\*油条\*, 那碗是弟弟吃饭专用的碗。  
盒子里装着\*一\*根\*小(大)\*麻花\*, 那盒子上写着十八街。  
课本里夹着\*一\*片\*小(大)\*树叶\*, 那本是语文课本。  
抽屉里放着\*一\*片\*小(大)\*光盘\*, 那抽屉还带着锁。  
窗台上摆着\*一\*片\*小(大)\*花瓣\*, 那窗台在朝阳的那面。  
桌子上放着\*一\*片\*小(大)\*羽毛\*, 那桌子是丽丽的梳妆台。  
床上放着\*一\*片\*小(大)\*尿布\*, 那床是给刚出生的弟弟的。  
包里装着\*一\*片\*小(大)\*面膜\*, 那包是从英国进口的。  
墙上挂着\*一\*块\*小(大)\*画板\*, 那墙是大卫刚刚粉刷过的。  
包里装着\*一\*块\*小(大)\*手绢\*, 那包是妈妈刚刚洗过的。  
盒子里装着\*一\*块\*小(大)\*玉佩\*, 那盒子是用绸缎裹着的。  
墙上突出了\*一\*块\*小(大)\*方砖\*, 那面墙就在奶奶家后院。  
沙发下掉着\*一\*块\*小(大)\*积木\*, 那沙发是从德国进口的。  
瓶子里装着\*一\*把\*小(大)\*石子\*, 那瓶子是肚大口小的造型。  
包里装着\*一\*把\*小(大)\*绳子\*, 那包是大卫的登山专用包。  
地板上掉着\*一\*根\*小(大)\*头发\*, 那地板是妈妈刚擦过的。  
院子里插着\*一\*根\*小(大)\*木棍\*, 那院子里经常聚集流浪猫狗。  
花坛里插着\*一\*根\*小(大)\*树枝\*, 那花坛是奶奶每天打理的。  
袋子里剩着\*一\*根\*小(大)\*香肠\*, 那袋子是从餐厅扔出来的。  
盘子里装着\*一\*片\*小(大)\*面包\*, 那盘子是用青花瓷做的。  
地板上掉着\*一\*片\*小(大)\*玻璃\*, 那地板是德国进口实木地板。  
盒子里放着\*一\*块\*小(大)\*肥皂\*, 那盒子是丽丽专门从国外买的。  
笔袋里装着\*一\*块\*小(大)\*橡皮\*, 那笔袋很旧颜色都褪了。  
盘子里盛着\*一\*块\*小(大)\*蛋糕\*, 那盘子是用塑料做的。  
袋子里剩着\*一\*块\*小(大)\*奶糖\*, 那袋子是从爷爷的包里翻出来的。

#### questions

盒子是日本买的  
镜子是两面镜  
衣架是美国进口的  
桌子是木质的  
碗是大家共用的  
地板是很旧的  
那面墙早上可以晒到太阳  
抽屉快坏了  
书包是奶奶给弟弟买的  
笔袋是丽丽自己买的  
冰箱是丽丽自己买的  
冰箱是大卫妈妈给他的  
盘子是妈妈用来放垃圾的  
盘子是给弟弟吃饭用的  
袋子是购物袋  
碗是弟弟专用的  
盒子上写着十八街  
那本课本是英语课本  
抽屉带着锁  
窗台朝阳  
桌子是丽丽的  
床是给弟弟婴儿床  
包是英国进口的  
墙是大卫新粉刷的  
包是奶奶刚买的  
盒子是木头漆盒  
墙在奶奶前院  
沙发是德国进口的  
瓶子的肚子很大  
那包是丽丽的化妆包  
地板是刚刚擦过的  
院子里经常聚集流浪猫狗  
花坛是爷爷在打理  
袋子是商场的环保袋  
盘子是青花瓷  
地板是德国进口的  
盒子是大卫送给丽丽的  
笔袋很新  
盘子是银质的  
袋子就挂在墙上

捕鼠夹上放着\*一\*块\*小（大）\*奶酪\*，那捕鼠夹就放在墙角桌子下。  
冰箱里放着\*一\*块\*小（大）\*黄油\*，那冰箱就立在墙角。

捕鼠夹在柜子里  
冰箱就立在墙角

## Condition 2: Num+Adj (small/big)+CL+Noun

### sentences

盒子里装着\*一\*小（大）\*把\*扇子\*，那盒子是平遥漆盒。  
镜子前摆着\*一\*小（大）\*把\*梳子\*，那镜子是两面镜。  
衣架上挂着\*一\*小（大）\*把\*伞\*，那衣架是从德国进口的。  
桌子上摆着\*一\*小（大）\*把\*叉子\*，那桌子是大理石做的。  
碗里放着\*一\*小（大）\*把\*勺子\*，那碗是专门给宝宝吃饭的。  
地板上掉着\*一\*小（大）\*把\*刀\*，那地板是新铺的实木地板。  
墙上挂着\*一\*小（大）\*把\*钥匙\*，那面墙只在下午能晒到太阳。  
抽屉里放着\*一\*小（大）\*把\*铜锁\*，那个抽屉很老几乎要坏掉了。  
书包里装着\*一\*小（大）\*把\*尺子\*，那书包是奶奶给弟弟买的。  
笔袋里装着\*一\*小（大）\*根\*铅笔\*，那笔袋是大卫送给丽丽的。  
冰箱里放着\*一\*小（大）\*根\*雪糕\*，那冰箱是丽丽从网上买的。  
冰箱里摆着\*一\*小（大）\*根\*萝卜\*，那冰箱是大卫新买的。  
盘子里放着\*一\*小（大）\*根\*黄瓜\*，那盘子是妈妈用来装瓜果的。  
盘子里放着\*一\*小（大）\*根\*骨头\*，那盘子是用来给狗狗喂食的。  
袋子里装着\*一\*小（大）\*根\*香蕉\*，那袋子是绿色环保的购物袋。  
碗里放着\*一\*小（大）\*根\*油条\*，那碗是弟弟吃饭专用的碗。  
盒子里装着\*一\*小（大）\*根\*麻花\*，那盒子上写着十八街。  
课本里夹着\*一\*小（大）\*片\*树叶\*，那本是语文课本。  
抽屉里放着\*一\*小（大）\*片\*光盘\*，那抽屉还带着锁。  
窗台上摆着\*一\*小（大）\*片\*花瓣\*，那窗台在朝阳的那面。  
桌子上放着\*一\*小（大）\*片\*羽毛\*，那桌子是丽丽的梳妆台。  
床上放着\*一\*小（大）\*片\*尿布\*，那床是给刚出生的弟弟的。  
包里装着\*一\*小（大）\*片\*面膜\*，那包是从英国进口的。  
墙上挂着\*一\*小（大）\*块\*画板\*，那墙是大卫刚刚粉刷过的。  
包里装着\*一\*小（大）\*块\*手绢\*，那包是妈妈刚刚洗过的。  
盒子里装着\*一\*小（大）\*块\*玉佩\*，那盒子是用绸缎裹着的。  
墙上突出了\*一\*小（大）\*块\*方砖\*，那面墙就在奶奶家后院。  
沙发下掉着\*一\*小（大）\*块\*积木\*，那沙发是从德国进口的。  
瓶子里装着\*一\*小（大）\*把\*石子\*，那瓶子是肚大口小的造型。  
包里装着\*一\*小（大）\*把\*绳子\*，那包是大卫的登山专用包。  
地板上掉着\*一\*小（大）\*根\*头发\*，那地板是妈妈刚擦过的。  
院子里插着\*一\*小（大）\*根\*木棍\*，那院子里经常聚集流浪猫狗。  
花坛里插着\*一\*小（大）\*根\*树枝\*，那花坛是奶奶每天打理的。  
袋子里剩着\*一\*小（大）\*根\*香肠\*，那袋子是从餐厅扔出来的。  
盘子里装着\*一\*小（大）\*片\*面包\*，那盘子是用青花瓷做的。  
地板上掉着\*一\*小（大）\*片\*玻璃\*，那地板是德国进口实木地板。  
盒子里放着\*一\*小（大）\*块\*肥皂\*，那盒子是丽丽专门从国外买的。  
笔袋里装着\*一\*小（大）\*块\*橡皮\*，那笔袋很旧颜色都褪了。  
盘子里盛着\*一\*小（大）\*块\*蛋糕\*，那盘子是用塑料做的。

### questions

盒子是日本买的  
镜子是两面镜  
衣架是美国进口的  
桌子是木质的  
碗是大家共用的  
地板是很旧的  
那面墙早上可以晒到太阳  
抽屉快坏了  
书包是奶奶给弟弟买的  
笔袋是丽丽自己买的  
冰箱是丽丽自己买的  
冰箱是大卫妈妈给他的  
盘子是妈妈用来放垃圾的  
盘子是给弟弟吃饭用的  
袋子是购物袋  
碗是弟弟专用的  
盒子上写着十八街  
那本课本是英语课本  
抽屉带着锁  
窗台朝阳  
桌子是丽丽的  
床是给弟弟婴儿床  
包是英国进口的  
墙是大卫新粉刷的  
包是奶奶刚买的  
盒子是木头漆盒  
墙在奶奶前院  
沙发是德国进口的  
瓶子的肚子很大  
那包是丽丽的化妆包  
地板是刚刚擦过的  
院子里经常聚集流浪猫狗  
花坛是爷爷在打理  
袋子是商场的环保袋  
盘子是青花瓷  
地板是德国进口的  
盒子是大卫送给丽丽的  
笔袋很新  
盘子是银质的

袋子里剩着\*一\*小(大)\*块\*奶糖\*, 那袋子是从爷爷的包里翻出来的。  
捕鼠夹上放着\*一\*小(大)\*块\*奶酪\*, 那捕鼠夹就放在墙角桌子下。  
冰箱里放着\*一\*小(大)\*块\*黄油\*, 那冰箱就立在墙角。

袋子就挂在墙上  
捕鼠夹在柜子里  
冰箱就立在墙角

### Condition 3: Num+Adj (small/big)+CL+de+Noun

#### sentences

盒子里装着\*一\*小(大)\*把\*的\*扇子\*, 那盒子是平遥漆盒。  
镜子前摆着\*一\*小(大)\*把\*的\*梳子\*, 那镜子是两面镜。  
衣架上挂着\*一\*小(大)\*把\*的\*伞\*, 那衣架是从德国进口的。  
桌子上摆着\*一\*小(大)\*把\*的\*叉子\*, 那桌子是大理石做的。  
碗里放着\*一\*小(大)\*把\*的\*勺子\*, 那碗是专门给宝宝吃饭的。  
地板上掉着\*一\*小(大)\*把\*的\*刀\*, 那地板是新铺的实木地板。  
墙上挂着\*一\*小(大)\*把\*的\*钥匙\*, 那面墙只在下午能晒到太阳。  
抽屉里放着\*一\*小(大)\*把\*的\*铜锁\*, 那个抽屉很老几乎要坏掉了。  
书包里装着\*一\*小(大)\*把\*的\*尺子\*, 那书包是奶奶给弟弟买的。  
笔袋里装着\*一\*小(大)\*根\*的\*铅笔\*, 那笔袋是大卫送给丽丽的。  
冰箱里放着\*一\*小(大)\*根\*的\*雪糕\*, 那冰箱是丽丽从网上买的。  
冰箱里摆着\*一\*小(大)\*根\*的\*萝卜\*, 那冰箱是大卫新买的。  
盘子里放着\*一\*小(大)\*根\*的\*黄瓜\*, 那盘子是妈妈用来装瓜果的。  
盘子里放着\*一\*小(大)\*根\*的\*骨头\*, 那盘子是用来给狗狗喂食的。  
袋子里装着\*一\*小(大)\*根\*的\*香蕉\*, 那袋子是绿色环保的购物袋。  
碗里放着\*一\*小(大)\*根\*的\*油条\*, 那碗是弟弟吃饭专用的碗。  
盒子里装着\*一\*小(大)\*根\*的\*麻花\*, 那盒子上写着十八街。  
课本里夹着\*一\*小(大)\*片\*的\*树叶\*, 那本是语文课本。  
抽屉里放着\*一\*小(大)\*片\*的\*光盘\*, 那抽屉还带着锁。  
窗台上摆着\*一\*小(大)\*片\*的\*花瓣\*, 那窗台在朝阳的那面。  
桌子上放着\*一\*小(大)\*片\*的\*羽毛\*, 那桌子是丽丽的梳妆台。  
床上放着\*一\*小(大)\*片\*的\*尿布\*, 那床是给刚出生的弟弟的。  
包里装着\*一\*小(大)\*片\*的\*面膜\*, 那包是从英国进口的。  
墙上挂着\*一\*小(大)\*块\*的\*画板\*, 那墙是大卫刚刚粉刷过的。  
包里装着\*一\*小(大)\*块\*的\*手绢\*, 那包是妈妈刚刚洗过的。  
盒子里装着\*一\*小(大)\*块\*的\*玉佩\*, 那盒子是用绸缎裹着的。  
墙上突出了\*一\*小(大)\*块\*的\*方砖\*, 那面墙就在奶奶家后院。  
沙发上掉着\*一\*小(大)\*块\*的\*积木\*, 那沙发是从德国进口的。  
瓶子里装着\*一\*小(大)\*把\*的\*石子\*, 那瓶子是肚大口小的造型。  
包里装着\*一\*小(大)\*把\*的\*绳子\*, 那包是大卫的登山专用包。  
地板上掉着\*一\*小(大)\*根\*的\*头发\*, 那地板是妈妈刚擦过的。  
院子里插着\*一\*小(大)\*根\*的\*木棍\*, 那院子里经常聚集流浪猫狗。  
花坛里插着\*一\*小(大)\*根\*的\*树枝\*, 那花坛是奶奶每天打理的。  
袋子里剩着\*一\*小(大)\*根\*的\*香肠\*, 那袋子是从餐厅扔出来的。  
盘子里装着\*一\*小(大)\*片\*的\*面包\*, 那盘子是用青花瓷做的。  
地板上掉着\*一\*小(大)\*片\*的\*玻璃\*, 那地板是德国进口实木地板。  
盒子里放着\*一\*小(大)\*块\*的\*肥皂\*, 那盒子是丽丽专门从国外买的。  
笔袋里装着\*一\*小(大)\*块\*的\*橡皮\*, 那笔袋很旧颜色都褪了。

#### questions

盒子是日本买的  
镜子是两面镜  
衣架是美国进口的  
桌子是木质的  
碗是大家共用的  
地板是很旧的  
那面墙早上可以晒到太阳  
抽屉快坏了  
书包是奶奶给弟弟买的  
笔袋是丽丽自己买的  
冰箱是丽丽自己买的  
冰箱是大卫妈妈给他的  
盘子是妈妈用来放垃圾的  
盘子是给弟弟吃饭用的  
袋子是购物袋  
碗是弟弟专用的  
盒子上写着十八街  
那本课本是英语课本  
抽屉带着锁  
窗台朝阳  
桌子是丽丽的  
床是给弟弟婴儿床  
包是英国进口的  
墙是大卫新粉刷的  
包是奶奶刚买的  
盒子是木头漆盒  
墙在奶奶前院  
沙发是德国进口的  
瓶子的肚子很大  
那包是丽丽的化妆包  
地板是刚刚擦过的  
院子里经常聚集流浪猫狗  
花坛是爷爷在打理  
袋子是商场的环保袋  
盘子是青花瓷  
地板是德国进口的  
盒子是大卫送给丽丽的  
笔袋很新

盘子里盛着\*一\*小（大）\*块\*的\*蛋糕\*，那盘子是用塑料做的。  
袋子里剩着\*一\*小（大）\*块\*的\*奶糖\*，那袋子是从爷爷的包里翻出来的。  
捕鼠夹上放着\*一\*小（大）\*块\*的\*奶酪\*，那捕鼠夹就放在墙角桌子下。  
冰箱里放着\*一\*小（大）\*块\*的\*黄油\*，那冰箱就立在墙角。

盘子是银质的  
袋子就挂在墙上  
捕鼠夹在柜子里  
冰箱就立在墙角



## Appendix B. Supplementary Results

### B.1. The best fit model results of L1ers' and L2ers' behaviour data in the Visual World Paradigm Experiment

Table.B.1. 1. L1ers'

	Estimate	Std. Error	Z-value	p-value
Nominal C2-C1	1.1987	0.4833	2.480	0.0131*
Nominal C3-C2	-2.8444	0.7916	-3.593	<0.001***
Nominal C3-C2	-4.0431	0.8122	-4.978	<0.001***
Noun M-C	-21.8908	10.786	-2.031	0.0266*
Classifier <i>gen-ba</i>	-6.3e-01	4.2e-01	-1.363	0.36
Classifier <i>kuai-ba</i>	-6.3e-01	4.2e-01	-1.363	0.36
Classifier <i>kuai-gen</i>	-2.9e-15	4.7e-01	0.001	0.99
Condition 2: Classifier <i>gen</i>	-1.5664	0.6935	-2.259	0.0238*
Condition 3: Classifier <i>gen</i>	2.1889	0.9470	2.311	0.0208*
Condition 2: Classifier <i>kuai</i>	-1.5624	0.6904	-2.261	0.0231*
Condition 3: Classifier <i>kuai</i>	2.1631	0.9187	2.345	0.0204*

Table.B.1. 2. L2ers'

	Estimate	Std. Error	Z-value	p-value
Nominal C2-C1	2.0711	0.4115	5.0330	<0.001***
Nominal C3-C2	-1.3972	0.6169	-2.2648	0.0149*
Nominal C3-C2	-3.5573	0.7144	-4.9794	<0.001***
Noun M-C	-19.1582	10.3174	-1.8568	0.0305*
Classifier <i>gen-ba</i>	-5.8e-03	4.2e-01	-0.0138	0.751
Classifier <i>kuai-ba</i>	-5.8e-01	4.2e-01	-1.380	0.618
Classifier <i>kuai-gen</i>	-2.9e-07	4.7e-01	-6.1E-07	0.99

Condition 2: Classifier <i>gen</i>	-1.2469	0.6188	-2.0150	0.0189*
Condition 3: Classifier <i>gen</i>	1.5976	0.8291	1.9269	0.0211*
Condition 2: Classifier <i>kuai</i>	-1.3428	0.6154	-2.1819	0.0164*
Condition 3: Classifier <i>kuai</i>	1.7998	0.9767	1.8427	0.0251*

### B.2. The results of Tukey post-hoc test on L1ers' and L2ers' behavioural data when the classifier was *ba* in the Visual World Paradigm Experiment

Table.B.2. 1. L1ers'

	Estimate	Standard Error	Z-value	p-value
Nominal C2-C1	1.198	0.4833	2.480	<b>0.0131*</b>
Nominal C3-C1	-2.844	0.7916	-3.593	<b>&lt;0.001***</b>
Nominal C3-C2	-4.043	0.8122	-4.978	<b>&lt;0.001***</b>
Noun M-C	-22.21	1.41	-15.762	<b>&lt;0.001***</b>

Table.B.2. 2. L2ers'

	Estimate	Standard Error	Z-value	p-value
Nominal C2-C1	1.277	0.6042	2.114	<b>0.0168*</b>
Nominal C3-C1	-2.895	0.6965	-4.1565	<b>&lt;0.001***</b>
Nominal C3-C2	-4.431	0.8203	-5.4016	<b>&lt;0.001***</b>
Noun M-C	-21.76	1.57	-13.859	<b>&lt;0.001***</b>

### B.3. The results of Tukey post-hoc test on L1ers' and L2ers' behavioural data when the classifier was *gen/kuai* in the Visual World Paradigm Experiment

Table.B.3. 1. L1ers'

	Estimate	Standard Error	Z-value	p-value
Gen: M-C	-20.3	7.14	-2.83	<b>&lt;0.001***</b>
Kuai: M-C	-20.62	6.81	-3.093	<b>&lt;0.001***</b>

Table.B.3. 2. L2ers'

	Estimate	Standard Error	Z-value	p-value
Gen: M-C	-18.03	8.74	-2.062	<0.001***
Kuai: M-C	-18.84	8.19	-2.31	<0.001***

#### B.4. The differences of total reading times on the phrase for L1ers in the Reading for Comprehension Experiment

	Estimate	Standard Error	df	t-value	p-value
Classifier:Nominal - <i>ba</i>					
Nominal t1 vs. t2	138.37	85.63	30.09	1.622	0.0504.
Nominal t1 vs. t3	205.97	109.47	29.92	1.881	0.0497*
Nominal t2 vs. t3	66.785	118.46	30.01	0.593	0.855
Classifier:Nominal - <i>gen</i>					
Nominal t1 vs. t2	76.87	135.95	32.94	0.565	0.57561
Nominal t1 vs. t3	403.47	136.01	33	2.966	0.00557**
Nominal t2 vs. t3	326.6	134.87	32.98	2.421	0.02111*
Classifier:Nominal - <i>kuai</i>					
Nominal t1 vs. t2	-69.55	153.65	30.13	-0.453	0.654
Nominal t1 vs. t3	239.13	153.65	30.13	1.556	0.031*
Nominal t2 vs. t3	308.68	153.25	29.94	2.106	0.029*
Classifier:Nominal - <i>pian</i>					
Nominal t1 vs. t2	-52.43	155.61	20.91	-0.337	0.74
Nominal t1 vs. t3	187.97	105.8	21.01	1.907	0.024*
Nominal t2 vs. t3	239.16	153.51	20.95	1.566	0.028*

#### B.5. The differences of first fixation durations on the classifier for L1ers and L2ers in the Reading for Comprehension Experiment

Table.B.5. 1. L1ers'

	Estimate	Standard Error	df	t-value	p-value
Classifier:Nominal - <i>ba</i>					
Nominal t1 vs. t2	67.67	29.09	31.32	2.326	0.0267*
Nominal t1 vs. t3	11.77	28.32	28.66	0.415	0.6809
Nominal t2 vs. t3	-55.91	29.06	31.58	-1.924	0.0634.
Classifier:Nominal - <i>gen</i>					
Nominal t1 vs. t2	5.887	27.381	37.84	0.181	0.822
Nominal t1 vs. t3	24.56	24.88	33.19	0.987	0.203
Nominal t2 vs. t3	19.752	27.326	38.29	0.723	0.474
Classifier:Nominal - <i>kuai</i>					
Nominal t1 vs. t2	-6.397	34.701	46.2	-0.178	0.932
Nominal t1 vs. t3	19.628	35.075	47.48	0.567	0.578
Nominal t2 vs. t3	25.55	27.67	28.34	0.928	0.207
Classifier:Nominal - <i>pian</i>					
Nominal t1 vs. t2	11.16	32.18	20.68	0.346	0.747
Nominal t1 vs. t3	-2.48	32.25	20.73	-0.076	0.925
Nominal t2 vs. t3	-14.47	31.67	20.28	-0.457	0.634

Table.B.5. 2. L2ers'

	Estimate	Std. Error	df	t-value	p-value
Classifier:Nominal - <i>ba</i>					
Nominal t1 vs. t2	273.23	99.09	30.08	2.757	0.00981**
Nominal t1 vs. t3	-54.76	98.69	30.59	-0.555	0.58299
Nominal t2 vs. t3	-328.01	105.63	29.91	3.101	0.004**
Classifier:Nominal - <i>gen</i>					
Nominal t1 vs. t2	180.7	69.9	157.3	2.58	0.01*
Nominal t1 vs. t3	104.8	73.3	146.6	1.43	0.15
Nominal t2 vs. t3	-75.9	77.7	63	-0.98	0.33
Classifier:Nominal - <i>kuai</i>					
Nominal t1 vs. t2	401.6	131.1	14	3.06	0.008**
Nominal t1 vs. t3	128.5	128.7	14.9	1	0.334

Nominal t2 vs. t3	-273.1	127.1	14.6	-2.15	0.049*
Classifier:Nominal - <i>pian</i>					
Nominal t1 vs. t2	111.43	119.31	16.38	0.934	0.364
Nominal t1 vs. t3	-46.57	116.15	15.68	-0.401	0.581
Nominal t2 vs. t3	-158	118.72	15.17	-1.331	0.203

#### **B.6. The differences of first fixation durations on nouns with *ba* for L1ers and L2ers in the Reading for Comprehension Experiment**

	Estimate	Std.Error	df	t value	Pr(> t )
Noun: count					
Nominal t1 vs. t2	40.96	19.21	22.69	2.132	0.0441*
Nominal t1 vs. t3	21.22	11.94	23.28	1.864	0.0498*
Nominal t2 vs. t3	-19.74	19.56	22.15	-1.009	0.3235
Noun: mass					
Nominal t1 vs. t2	-38.76	20.06	2.68	-1.935	0.0438*
Nominal t1 vs. t3	-80.41	27.46	2.68	-2.979	0.0267*
Nominal t2 vs. t3	-41.64	28.79	2.27	-1.412	0.0909

#### **B.7. The differences of first fixation durations on nouns for L2ers in the Reading for Comprehension Experiment**

	Estimate	Standard Error	df	t-value	p-value
Ba: c - m	-1.4	96.35	100.8	-0.01	0.988
Gen: c - m	-153.4	75.72	104.9	-2.03	0.045*
Kuai: c - m	-98.7	50.41	105.9	-1.96	0.062.
Pian: c - m	-278.1	103.46	114.7	-2.69	0.008**

## B.8. The differences of total reading times on the nouns for L1ers and L2ers in the Reading for Comprehension Experiment

Table.B.8. 1. L1ers'

	Estimate	Std. Error	df	t value	Pr(> t )
Ba vs. Gen	128.24	43.73	116.41	2.926	0.0165*
Ba vs. Kuai	84.34	40.57	118.65	2.138	0.0204*
Ba vs. Pian	143.38	47.55	117.47	3.045	0.0107*
Gen vs. Kuai	-44.16	40.08	116.38	1.108	0.2961
Gen vs. Pian	15.51	35.84	116.48	0.433	0.6659
Kuai vs. Pian	59.36	39.43	116.27	1.501	0.2321

Table.B.8. 2. L2ers'

	Estimate	Standard Error	df	t-value	p-value
Ba vs. Gen	619	135	118	4.57	<2e-16***
Ba vs. Kuai	353	141	122	2.51	0.01*
Ba vs. Pian	511	152	122	3.36	0.001***
Gen vs. Kuai	-266	153	117	-1.74	0.0987
Gen vs. Pian	-108	146	116	-0.74	0.461
Kuai vs. Pian	246	149	118	1.64	0.103

## References

- Ahrens, K. (1994). Classifier production in normals and aphasics. *Journal of Chinese Linguistics*, 22(2), 202-247.
- Allan, K. (1977). Classifiers. *Language*, 53, 285-311.
- Alloppenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, 38: 419–439.
- Allport, D., Antonis, B., & Reynolds, P. (1972). On the division of attention: A disproof of the single channel hypothesis. *Quarterly Journal of Experimental Psychology*, 24, 225-235.
- Altmann, G. T. M. (2004). Language-mediated eye movements in the absence of a visual world: The ‘blank screen paradigm’. *Cognition*, 93, 79–87.
- Altmann, G. T. M., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language*, 57, 502–518.
- Altmann, G., & Mirković, J. (2009). Incrementality and prediction in human sentence processing. *Cognitive science*, 33(4), 583-609.
- Baddeley, A. D. (1999). *Human memory*. Boston: Allyn & Bacon.
- Baltova, I., 1999. Multisensory Language Teaching in a Multidimensional Curriculum: The Use of Authentic Bimodal Video in Core French. *Canadian Modern Language Review*, 56 (1), 32-48.
- Barner, D., S. Inagaki., and P. Li. (2009). Language, though, and real nouns. *Cognition*, 111 (3): 329-344.
- Barr, D. J., Gann, T. M., and Pierce, R. S. (2011). Anticipatory baseline effects and information integration in visual world studies. *Acta Psychologica*, 137: 201–07.

- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, 68(3), 255-278.
- Bates, D., Maechler, M., & Dai, B. (2008). The lme4 package. *Computer software manual*]. Retrieved from <http://cran.r-project.org/web/packages/lme4/lme4.pdf>.
- Bates, E., & MacWhinney, B. (1989). Functionalism and the Competition Model. In B. MacWhinney & E. Bates (eds.), *The cross-linguistic study of sentence processing*, pp. 3–73. Cambridge: Cambridge University Press.
- Benmamoun, E., Montrul, S., & Polinsky, M. (2013a). Heritage languages and their speakers: Opportunities and challenges for linguistics. *Theoretical Linguistics*, 39(3-4), 129-181.
- Benmamoun, E., Montrul, S., & Polinsky, M. (2013b). Defining an “ideal” heritage speaker: Theoretical and methodological challenges Reply to peer commentaries. *Theoretical Linguistics*, 39(3-4), 259-294.
- Blair, I. V., Urland, G. R., & Ma, J. E. (2002). Using Internet search engines to estimate word frequency. *Behavior Research Methods, Instruments, & Computers*. 34, 286- 290.
- Bley-Vroman, R. (1989). What is the logical problem of foreign language learning? In S. M. Gass & J. Schachter (Eds.), *Linguistic perspectives on second language acquisition* (pp. 1–68). Cambridge, UK: Cambridge University Press.
- Bley-Vroman, R. (2009). The evolving context of the fundamental difference hypothesis. *Studies in Second Language Acquisition*, 31, 175–198
- Blumenfeld, H. K., & Marian, V. (2011). Bilingualism influences inhibitory control in auditory comprehension. *Cognition*, 118, 245– 257.
- Borer, H. (2005). *In name only* (Vol. 1). Oxford University Press on Demand.
- Brown, R. (1973). *A first language/The early stages*. Cambridge, MA: Harvard University Press.
- Buswell, G. T. (1935). How people look at pictures: a study of the psychology and perception in art.
- Cai, Z. G., Pickering, M. J., & Branigan, H. P. (2012). Mapping concepts to syntax:



- Evidence from structural priming in Mandarin Chinese. *Journal of Memory and Language*, 66(4), 833-849.
- Carey, S. (1978). The child as word learner. In M. Halle, J. Bresnan and G. A. Miller (Eds.) *Linguistic theory and psychological reality* (pp. 264-293). Cambridge: The MIT press.
- Chao, Y. R. (1968). *A grammar of spoken Chinese*. University of California Press, Berkeley.
- Chang, H. W. (1983). *Pre-schooler's use of classifiers in mandarin Chinese*. National Taiwan University Press.
- Chen, L., Shu, H., Liu, Y. Y., & Zhao, J. J. (2007). ERP signatures of subject-verb agreement in L2 learning. *Bilingualism: Language and Cognition*, 10(2), 161-174.
- Cheng, L. L. S., & Sybesma, R. (1998). Yi-wan tang, yi-ge tang: Classifiers and massifiers. *Tsing Hua journal of Chinese studies*, 28(3), 385-412.
- Cheng, L. L. S., & Sybesma, R. (1999). Bare and not-so-bare nouns and the structure of NP. *Linguistic inquiry*, 30(4), 509-542.
- Cheng, L. L. S., Doetjes, J., & Sybesma, R. (2008). How universal is the Universal Grinder. *Linguistics in the Netherlands*, 2008, 50-62.
- Cheng, L. L. S., & Sybesma, R. (2012). Classifiers and DP. *Linguistic inquiry*, 43(4), 634-650.
- Chierchia, G. (1994). Syntactic bootstrapping and the acquisition of noun meanings: The mass-count issue. *Syntactic theory and first language acquisition: Crosslinguistic perspectives*, 1, 301-318.
- Chierchia, G. (1998a). *Plurality of mass nouns and the notion of "semantic parameter"* (pp. 53-103). Springer Netherlands.
- Chierchia, G. (1998a). *Plurality of mass nouns and the notion of "semantic parameter"* (pp. 53-103). Springer Netherlands.
- Chierchia, G. (1998b). Reference to kinds across language. *Natural language semantics*, 6(4), 339-405.
- Chierchia, G. (2010). Mass nouns, vagueness and semantic variation. *Synthese*, 174(1), 99-149.

- Chien, Y. C., Lust, B., & Chiang, C. P. (2003). Chinese children's comprehension of count-classifiers and mass-classifiers. *Journal of East Asian Linguistics*, 12, 91-120.
- Choi, S.H., Ionin, T., and Zhu, Y. Q. (2017). L1 Korean and L1 Mandarin L2 English learners' acquisition of the count/mass distinction in English. *Second language research*, July, 1-31.
- Chou, C. J., Huang, H. W., Lee, C. L., & Lee, C. Y. (2014). Effects of semantic constraint and cloze probability on Chinese classifier-noun agreement. *Journal of Neurolinguistics*, 31, 42-54.
- Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive psychology*, 42(4), 368-407.
- Chu, K. H. (2015). Activation of grammatical representations during the reading of Chinese compound words. HKU Theses Online (HKUTO).
- Clahsen, H., & Felser, C. (2006a). Grammatical processing in first and second language learners. *Applied Psycholinguistics*, 27, 3-42.
- Clahsen, H., & Felser, C. (2006b). How native-like is non-native language processing? *Trends in cognitive sciences*, 10(12), 564-570.
- Clahsen, H., & Felser, C. (2006c). Continuity and shallow structures in language processing. *Applied Psycholinguistics*, 27(01), 107-126.
- Cleary, A. M., & Langley, M. M. (2007). Retention of the structure underlying sentences. *Language and Cognitive Processes*, 22, 614-628.
- Clifton, C., Ferreira, F., Henderson, J. M., Inhoff, A. W., Liversedge, S. P., Reichle, E. D., & Schotter, E. R. (2016). Eye movements in reading and information processing: Keith Rayner's 40year legacy. *Journal of Memory and Language*, 86, 1-19.
- Clifton, C., Staub, A., & Rayner, K. (2007). Eye movements in reading words and sentences. In R. P. G. Van Gompel, M. H. Fischer, W. S. Murray, & R. L. Hill (Eds.), *Eye-movements: A window on mind and brain* (pp. 241-372). Oxford: Elsevier Science.

- Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology*, 6, 84-107.
- Croft, W. (1994). Semantic universals in classifier systems. *Word*, 45, 145–171.
- Cutler, A. (1983). Speakers' conceptions of the function of prosody. In *Prosody: Models and measurements* (pp. 79-91). Springer Berlin Heidelberg.
- DeKeyser, R. M. (2000). The robustness of critical period effects in second language acquisition. *Studies in second language acquisition*, 22(4), 499-533.
- DeKeyser, R. M. (2005). What makes learning second-language grammar difficult? A review of issues. *Language Learning*, 55 (S1), 1–25.
- Dekydtspotter, L., Schwartz, B. D., & Sprouse, R. A. (2006). The comparative fallacy in L2 processing research. In *Proceedings of the 8th generative approaches to second language acquisition conference (GASLA 2006): The Banff conference* (pp. 33-40).
- Doetjes, J. (1997). Quantifiers and selection. On the distribution of quantifying expressions in French, Dutch and English, The Hague: Holland Academic Graphics (HIL dissertation series).
- Donaldson, B. (2011). Left dislocation in near-native French. *Studies in Second Language Acquisition*, 33(3), 399-432.
- Duffy, S. A., Morris, R. K., & Rayner, K. (1988). Lexical ambiguity and fixation times in reading. *Journal of Memory and Language*, 27, 429–446.
- Duyck, W., Van Assche, E., Drieghe, D., & Hartsuiker, R. J. (2007). Visual word recognition by bilinguals in a sentence context: Evidence for non-selective lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33, 663–679.
- Eberhard, K. M., Spivey-Knowlton, M. J., Sedivy, J. C., & Tanenhaus, M. K. (1995). Eye movements as a window into real-time spoken language comprehension in natural contexts. *Journal of psycholinguistic research*, 24(6), 409-436.
- Ehrlich, S. F., & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning and Verbal Behavior*, 20,

641–655.

- Ellis, N. C. (1994). *Implicit and explicit learning of languages*. New York, NY: Academic Press.
- Ellis, N. C. (2005). At the interface: Dynamic interactions of explicit and implicit language knowledge. *Studies in Second Language Acquisition*, 27, 305–352.
- Ellis, N. C., Hafeez, K., Martin, K. I., Chen, L., Boland, J., & Sagarra, N. (2014). An eye-tracking study of learned attention in second language acquisition. *Applied Psycholinguistics*, 35(03), 547–579.
- Ellis, N. C., & Sagarra, N. (2011). Learned attention in adult language acquisition. *Studies in Second Language Acquisition*, 33(04), 589–624.
- Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in Second Language Acquisition*, 27, 141–172.
- Fang, F. X. (1985). An experiment on the use of classifiers by 4- to 6-year olds. *Acta Psychologica Sinica*, 17(4): 348–392.
- Federmeier, K. D., Wlotko, E. W., De Ochoa-Dewald, E., & Kutas, M. (2007). Multiple effects of sentential constraint on word processing. *Brain Research*, 1146, 75–84.
- Felser, C., Roberts, L., Marinis, T., & Gross, R. (2003). The processing of ambiguous sentences by first and second language learners of English. *Applied Psycholinguistics*, 24(3), 453.
- Felser, C., Sato, M., & Bertenshaw, N. (2009). The on-line application of binding Principle A in English as a second language. *Bilingualism: Language and Cognition*, 12, 485–502.
- Flecken, M. (2011). Event conceptualization by early Dutch-German bilinguals: Insights from linguistic and eye-tracking data. *Bilingualism: Language and Cognition*, 14, 61–77.
- Foote, R. (2011). Integrated knowledge of agreement in early and late English-Spanish bilinguals. *Applied Psycholinguistics*, 32, 187–220.
- Foucart, A., & Frenck-Mestre, C. (2012). Can late L2 learners acquire new grammatical features? Evidence from ERPs and eye-tracking. *Journal of Memory and Language*, 66(1), 226–248.

- Foucart, A., Martin, C. D., Moreno, E. M., & Costa, A. (2014). Can bilinguals see it coming? Word anticipation in L2 sentence reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40(5), 1461.
- Franceschina, F. (2001). Morphological or syntactic deficits in near-native speakers? An assessment of some current proposals. *Second Language Research*, 17(3), 213-247.
- Frenck-Mestre, C. (2002). An on-line look at sentence processing in the second language. In Heredia, R. and Altarriba, J., editors, *Bilingual sentence processing*. Elsevier, 217–36.
- Frenck-Mestre, C. (2005). Eye-movement recording as a tool for studying syntactic processing in a second language: A review of methodologies and experimental findings. *Second Language Research*, 21(2), 175-198.
- Frenck-Mestre, C., Foucart, A., Carrasco-Ortiz, H., & Herschensohn, J. (2009). Processing of grammatical gender in French as a first and second language: Evidence from ERPs. *Eurosla yearbook*, 9(1), 76-106.
- Frick, R. (1984). Using both an auditory and a visual short-term store to increase digit span. *Memory & Cognition*, 12, 507-514.
- Gao, H. H. (2009). Cognitive barriers to the learning of Chinese noun classifiers by native Swedish speakers. Paper presented at Second conference of the Swedish association for language and cognition, Stockholm University.
- Gao, M., Malt, B. (2009). Mental representation and cognitive consequences of Chinese individual classifiers. *Language and Cognitive Processes*, 24(7-8): 1124-1179.
- Gass, S. M., & Selinker, L. (eds.). (1992). *Language transfer in language learning* (revised edition). Amsterdam: John Benjamins.
- Ginns, P. (2005). Meta-analysis of the modality effect. *Learning and Instruction*, 15, 313–331.
- Godfroid, A., & Winke, P. (2015). Investigating implicit and explicit processing using L2 learners' eye-movement data. *Implicit and explicit learning of languages*, 325-348.

- Godfroid, A., Winke, P., & Gass, S. (Eds.). (2013). Thematic issue on eye tracking in second language acquisition research. *Studies in Second Language Acquisition*, 35(2).
- Gong, J. S. (2010). Chinese classifier acquisition: comparison of L1 child and L2 adult development. *Theses, Dissertations, Professional Papers*: 192.
- Grüter, T., Lew-Williams, C., & Fernald, A. (2012). Grammatical gender in L2: A production or a real-time processing problem? *Second Language Research*, 28(2), 191-215.
- Grüter, T., & Rohde, H. (2013). L2 processing is affected by RAGE: Evidence from reference resolution. *In the 12th conference on Generative Approaches to Second Language Acquisition (GASLA)*.
- Gu, Y. (2013). Second language vocabulary. In J. Hattie, & E. M. Anderson (Eds.), *International guide to student achievement* (pp. 307–309). New York: Routledge.
- Guichon, N., & McLornan, S. (2008). The effects of multimodality on L2 learners: Implications for CALL resource design. *System*, 36(1), 85-93.
- Hahne, A., & Friederici, A. D. (2001). Processing a second language: Late learners' comprehension mechanisms as revealed by event-related brain potentials. *Bilingualism: Language and Cognition*, 4, 123–141.
- Hahne, A., Mueller, J. L., & Clahsen, H. (2006). Morphological processing in a second language: Behavioral and event-related brain potential evidence for storage and decomposition. *Journal of Cognitive Neuroscience*, 18, 121–134.
- Hawkins, R. (2001). *Second Language Syntax*. Blackwell
- Hawkins, R. (2009). Statistical learning and innate knowledge in the development of second language proficiency: Evidence from the acquisition of gender concord. In: Benati AG (ed.) *Issues in Second Language Proficiency*. London: Continuum, 63–78.
- Hawkins, R., & Hattori, H. (2006). Interpretation of English multiple wh-questions by Japanese speakers: A missing uninterpretable feature account. *Second Language Research*, 22, 269–301.

- Hawkins, R., & Liszka, S. (2003). Locating the source of defective past tense marking in advanced L2 English speakers. *Language Acquisition and Language Disorders*, 30, 21-44.
- Haznedar, B., & Schwartz, B. D. (1997). Are there optional infinitives in child L2 acquisition. In *Proceedings of the 21st annual Boston University conference on language development* (Vol. 21, pp. 257-68).
- Herrlinger, S., Höffler, T. N., Opfermann, M., & Leutner, D. (2016). When Do Pictures Help Learning from Expository Text? Multimedia and Modality Effects in Primary Schools. *Research in Science Education*, 1-20.
- Hill, R. L., & Murray, W. S. (2000). Chapter 22: commas and spaces: effects of punctuation on eye movements and sentence parsing. *Reading as a Perceptual Process*. A. Kennedy, R. Radach, D. Heller and J. Pynte (Editors).
- Hinkel, E. (2013). Language teaching curricula. In J. Hattie, & E. M. Anderson (Eds.), *International guide to student achievement* (pp. 310–311). New York: Routledge.
- Hogaboam, T. W., & Perfetti, C. A. (1975). Lexical ambiguity and sentence comprehension. *Journal of verbal learning and verbal behavior*, 14(3), 265-274.
- Hopp, H. (2006). Syntactic features and reanalysis in near-native processing. *Second Language Research*, 22 (3), 369-397.
- Hopp, H. (2010). Ultimate attainment in L2 inflectional morphology: Performance similarities between non-native and native speakers. *Lingua*, 120, 901–31.
- Hopp, H. (2012). The on-line integration of inflection in L2 processing: Predictive processing of German gender. In A. K. Biller, E. Y. Chung & A. E. Kimball (eds.), *Proceedings of the 36th Boston University Conference on Language Development* (BUCLD 36), pp. 226–245. Somerville, MA: Cascadilla Press.
- Hopp, H. (2015a). Differential learning of prediction. *Linguistic Approaches to Bilingualism*, 5, 476–81.
- Hopp, H. (2015b). Semantics and morphosyntax in L2 predictive sentence processing. *International Review of Applied Linguistics*, 53, 277–306.
- Hopp, H. (2016). Learning (not) to predict: Grammatical gender processing in second language acquisition. *Second Language Research*, 32(2), 277-307.

- Hsieh, Miao-Ling. (2008). *The Internal Structure of Noun Phrases in Chinese*. Taipei: Crane.
- Hu. Q. (1993a). *The acquisition of Chinese classifiers by young Mandarin-speaking children*. Ph.D. dissertation. Boston University.
- Hu. Q. (1993b). Overextension of animacy in Chinese classifier acquisition. In E. Clark (ed.). *The proceedings of the twenty-fifth annual child language research forum, centre for the study of language and information*, Stanford, California, 127-136.
- Hua Yan, Linying Liu, Michelle Tate, Lisa Wang, Yu Bin and Xiaoming Zhu. (2009). *Edexcel GCSE Chinese*. United Kingdom: Pearson Education Limited.
- Huetting, F., & Altmann, G. T. (2005). Word meaning and the control of eye fixation: Semantic competitor effects and the visual world paradigm. *Cognition*, 96(1), B23-B32.
- Huetting, F., Cheng, J. D., Bowerman, M., & Majid, A. (2010). Do language-specific categories shape conceptual processing? Mandarin classifier distinctions influence eye gaze behaviour, but only during linguistic processing. *Journal of Cognition and Culture*, 10: 39-58.
- Huetting, F. and McQueen, J. M. (2007). The tug of war between phonological, semantic, and shape information in language-mediated visual search. *Journal of Memory and Language*, (54): 460-482.
- Huetting, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, 137: 151-171.
- Hulstijn, J. H. (2005). Theoretical and empirical issues in the study of implicit and explicit second-language learning: Introduction. *Studies in second language acquisition*, 27(02), 129-140.
- Imai, M., Saalbach, H., & Stern, E. (2010). Are Chinese and German children taxonomic, thematic, or shape biased? Influence of classifiers and cultural contexts. *Frontiers in psychology*, 1, 194.
- Inagaki, S. (2014). Syntax-semantics mappings as a source of difficulty in Japanese speakers' acquisition of the mass-count distinction in English. *Bilingualism: Language and cognition*, 17 (3): 464-477.



- Inhoff, A. W., & Rayner, K. (1986). Parafoveal word processing during eye fixations in reading: Effects of word frequency. *Perception & Psychophysics*, 40 (6), 431-439.
- Ionin, T., Baek, S., Kim, E., Ko, H., & Wexler, K. (2012). That's not so different from the: definite and demonstrative descriptions in second language acquisition. *Second language research*, 28 (1), 69-101.
- Ionin, T., Ko, H., & Wexler, K. (2004). Article semantics in L2 acquisition: The role of specificity. *Language Acquisition*, 12(1), 3-69.
- Ionin, T., & Montrul, S. (2009). Article use and generic reference: Parallels between L1- and L2-acquisition. In M. Garc'ia-Mayo & R. Hawkins (Eds.), *Second language acquisition of articles: Empirical findings and theoretical implications* (pp. 147–173). Amsterdam: Benjamins.
- Ionin, T., & Montrul, S. (2010). The role of L1 transfer in the interpretation of articles with definite plurals in L2 English. *Language learning*, 60:4, 877-925.
- Ionin, T., Montrul, S., & Crivos, M. (2013). A bidirectional study on the acquisition of plural noun phrase interpretation in English and Spanish. *Applied Psycholinguistics*, 34, 483-518.
- Irwin, D. E. (2004). Fixation location and fixation duration as indices of cognitive processing. In J. Henderson & F. Ferreira (Eds.), *The interface of language, vision, and action: Eye movements and the visual world* (pp. 105–134). New York: Psychology Press.
- Jackson, C. N. (2007). The use and non-use of semantic information, word order, and case marking during comprehension by L2 learners of German. *Modern Language Journal*, 91, 418–432.
- Jackson, C. N., & Dussias, P. E. (2009). Cross-linguistic differences and their impact on L2 sentence processing. *Bilingualism: Language and Cognition*, 12, 65–82.
- Jarvis, S. (2002). Topic continuity in L2 English article use. *Studies in Second Language Acquisition*, 24, 387–418.
- Jiang, N. (2004). Morphological insensitivity in second language processing. *Applied Psycholinguistics*, 25, 603–634.

- Jiang, N. (2007). Selective integration of linguistic knowledge in adult second language acquisition. *Language Learning*, 57, 1–33.
- Jiang, N., Novokshanova, E., Masuda, K., & Wang, X. (2011). Morphological congruency and the acquisition of L2 morphemes. *Language learning*, 61(3), 940-967.
- Juffs, A., & Harrington, M. (2011). Aspects of working memory in L2 learning. *Language Teaching*, 44(02), 137-166.
- Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87(4), 329–354.
- Kaan, E. (2014). Predictive sentence processing in L2 and L1: What is different? *Linguistic Approaches to Bilingualism*, 4(2), 257-282.
- Kaan, E., Dallas, A. C., & Wijnen, F. (2010). Syntactic predictions in second-language sentence processing. J.-W. Zwart & M. de Vries (eds.), *Structure preserved. Festschrift in the honor of Jan Koster*, 207-213.
- Kamide Y (2008) Anticipatory processes in sentence processing. *Language and Linguistics Compass*, (2), 647–70.
- Kanero, J., Imai, M., Okada, H., & Hoshino, N. (2015). Do classifiers make the syntactic count/mass distinction? Insights from ERPs in classifier processing in Japanese. *Journal of Memory and Language*, 83, 20-52.
- Keating, G.D. (2009). Sensitivity to violations of gender agreement in native and non-native Spanish: An eye-movement investigation. *Language Learning*, 59(3), 503–535.
- Keating, G.D. (2013). Eye-tracking with text. In J. Jegerski & B. VanPatten (Eds.), *Research methods in second language psycholinguistics* (pp. 69–92). New York, NY: Routledge.
- Klein, N. M., Carlson, G. N., Li, R. J., Jaeger, T. F., & Tanenhaus, M. K. (2012). Classifying and massifying incrementally in Chinese language comprehension. *Count and mass across languages*, 261-282.

- Kliegl, R., Nuthmann, A., & Engbert, R. (2006). Tracking the mind during reading: The influence of past, present, and future words on fixation durations. *Journal of Experimental Psychology: General*, 135, 12–35.
- Ko, H., Ionin, T., & Wexler, K. (2010). The role of presuppositionality in the second language acquisition of English articles. *Linguistic Inquiry*, 41, 213–254.
- Kormos, J., & Sáfár, A. (2008). Phonological short-term memory, working memory and foreign language performance in intensive language learning. *Bilingualism: Language and cognition*, 11(02), 261-271.
- Kotz, S. A. (2008). A critical review of ERP and fMRI evidence on L2 syntactic processing. *Brain and Language*, 109, 68–74.
- Kotz, S. A., Holcomb, P. J., & Osterhout, L. (2008). ERPs reveal comparable syntactic sentence processing in native and non-native readers of English. *Acta Psychologica*, 128, 514–527.
- Köhne, J., & Crocker, M. W. (2010). Sentence processing mechanisms influence cross-situational word learning. In *Proceedings of the Annual Conference of the Cognitive Science Society*.
- Krashen, S. (1994). *The input hypothesis and its rivals. Implicit and explicit learning of languages*, 45-77.
- Krifka, M. (1995). Common nouns: A contrastive analysis of Chinese and English. In J. Pelletier & G. N. Carlson (Eds.). *The generic book*. University of Chicago Press.
- Lakoff, G. (1986). Classifiers as a reflection of mind. In C. Craig (Ed.), *Noun classes and categorization* (pp. 13-51). Philadelphia: John Benjamins.
- Lai, M. L., Tsai, M. J., Yang, F. Y., Hsu, C. Y., Liu, T. C., Lee, S. W. Y., ... & Tsai, C. C. (2013). A review of using eye-tracking technology in exploring learning from 2000 to 2012. *Educational Research Review*, 10, 90-115.
- Land, M. F., Mennie, N., & Rusted, J. (1999). Eye movements and the roles of vision in activities of daily living: making a cup of tea. *Perception*, 28, 1311–1328.
- Lardiere, D. (2000). Mapping features to forms in second language acquisition. *Second language acquisition and linguistic theory*, 102-129.

- Lardiere, D. (2009). Some thoughts on the contrastive analysis of features in second language acquisition. *Second language research*, 25 (2): 173-227.
- Lau, E., & Grüter, T. (2015). Real-time Processing of Classifier Information by L2 Speakers of Chinese. *BUCLD 39 Proceedings*, Cascadilla Press.
- Leal, T., Slabakova, R., & Farmer, T. A. (2017). The fine-tuning of linguistic expectations over the course of L2 learning. *Studies in Second Language Acquisition*, 39(3), 493-525.
- Lenneberg, E. H. (1967). The biological foundations of language. *Hospital Practice*, 2(12), 59-67.
- Leung, J. H. C., & Williams, J. N. (2012). Constraints on implicit learning of grammatical form-meaning connections. *Language Learning*, 62, 634–662.
- Leung, J. H., & Williams, J. N. (2014). Cross-linguistic differences in implicit language learning. *Studies in second language acquisition*, 36(04), 733-755.
- Levie, W. H., & Lentz, R. (1982). Effects of text illustrations—a review of research. *Educational Communication and Technology Journal*, 30, 195–232.
- Levin, J. R., Anglin, G. J., & Carney, R. N. (1987). On empirically validating functions of pictures in prose. In D. M. Willows & H. A. Houghton (Eds.), *The psychology of illustration: I. Basic research* (pp. 51–85). New York: Springer.
- Lew-Williams, C., & Fernald, A. (2009). Fluency in using morphosyntactic cues to establish reference: How do native and non-native speakers differ. In *Proceedings of the 33rd annual Boston university conference on language development*, Vol. 1, 290-301.
- Lew-Williams, C., & Fernald, A. (2010). Real-time processing of gender-marked articles by native and non-native Spanish speakers. *Journal of Memory and Language*, 63(4), 447-464.
- Lew-Williams, C., & Fernald, A. (2010). Real-time processing of gender-marked articles by native and non-native Spanish speakers. *Journal of Memory and Language*, 63, 447–464.

- Li, P., Barner, D., & Huang, B. (2008). Classifiers as count syntax: Individuation and measurement in the acquisition of Mandarin Chinese. *Language Learning and Development*, 4(4): 1-42.
- Li, X., & Rothstein, S. (2012). Measure Readings of Mandarin Classifier Phrases and the Particle de\*. *Language and Linguistics*, 13(4), 693.
- Li, X. S., Liu, P. P., & Rayner, K. (2011). Eye movement guidance in Chinese reading: Is there a preferred viewing location? *Vision Research*, 51: 1146-1156.
- Liang, S. Y. (2009). The acquisition of Chinese nominal classifiers by L2 adult learners. In Majorie K. M. Chan & Hana Kang (Eds.), *Proceedings of the 20<sup>th</sup> North American Conference on Chinese Linguistics (NACCL-20)* (pp. 309-326). Columbus, OH: Ohio State University.
- Libben, G. (2006). Getting at psychological reality: On- and off-line tasks in the investigation of hierarchical morphological structure. In G. Wiebe, G. Libben, T. Priestly, R. Smyth, & H. S. Wang (Eds.), *Phonology, morphology, and the empirical imperative: Papers in honour of Bruce L. Derwing* (pp. 235-255). Taipei: Crane Publishing.
- Lim, J. H., & Christianson, K. (2015). Second language sensitivity to agreement errors: Evidence from eye movements during comprehension and translation. *Applied Psycholinguistics*, 36(06), 1283-1315.
- Lin, L., & Atkinson, R. K. (2011). Using animations and visual cueing to support learning of scientific concepts and processes. *Computers & Education*, 56(3), 650–658.
- Liversedge, S. P., Zang, C., Zhang, M., Bai, X. J., Yan, G. L., & Drieghe D. (2014). The effect of visual complexity and word frequency on eye movements during Chinese reading. *Visual Cognition*, 22: 3-4, 441-457.
- Lucy, J. A. (1992). *Grammatical categories and cognition: A case study of the linguistic relativity hypothesis*. Cambridge: Cambridge University Press.
- Luk, Z. P., & Shirai, Y. (2009). Is the acquisition order of grammatical morphemes impervious to L1 knowledge? Evidence from the acquisition of plural -s, articles and possessive's. *Language Learning*, 59, 721–754.

- MacWhinney, B. (1987). *The competition model. Mechanisms of language acquisition*, 249-308.
- MacWhinney, B. (2001). The competition model: The input, the context, and the brain. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 69–90). New York, NY: Cambridge University Press.
- MacWhinney, B. (2005). A unified model of language acquisition. In J. F. Kroll & A. M. B. de Groot (eds.), *Handbook of bilingualism: Psycholinguistic approaches*, pp. 49–67. Oxford: Oxford University Press.
- MacWhinney, B. (2008). A unified model. In N. Ellis & P. Robinson (Eds.), *Handbook of cognitive linguistics and second language acquisition* (pp. 341–371). London, UK: Routledge.
- Malpass, D. & Meyer, A. S. (2010). Parallel processing of objects in a naming task: Evidence from extrafoveal-on-foveal effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.
- Marian, V., & Spivey, M. (2003). Bilingual and monolingual processing of competing lexical items. *Applied Psycholinguistics*, 24, 173–193.
- Marian, V., Spivey, M., & Hirsch, J. (2003). Shared and separate systems in bilingual language processing: Converging evidence from eye-tracking and brain imaging. *Brain and Language*, 86, 70–82.
- Mayer, R. E. (Ed.). (2005). *The Cambridge handbook of multimedia learning*. Cambridge: Cambridge University.
- Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). New York, NY: Cambridge University Press.
- Mayer, R. (2010). Unique contribution of eye-tracking research to the study of learning with graphics. *Learning and Instruction*, 20, 167–171.
- Mayer, R. E., & Fiorella, L. (2014). Principles for reducing extraneous cognitive processing in multimedia learning: Coherence, signalling, redundancy, spatial continuity, and temporal contiguity principles. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd edn; pp. 279–315). New York: Cambridge University Press.

- Mayer, R. E., & Johnson, C. I. (2008). Revising the redundancy principle in multimedia learning. *Journal of Educational Psychology*, 100, 380–386.
- Mayer, R. E., Heiser, H., & Lonn, S. (2001). Cognitive constraints on multimedia learning: When presenting more material results in less understanding. *Journal of Educational Psychology*, 93, 187–198.
- Mayer, R. E., Lee, H., & Peebles, A. (2014). Multimedia learning in a second language: A Cognitive load perspective. *Applied Cognitive Psychology*, 28(5), 653-660.
- McDonald, J. (2000). Grammaticality judgments in a second language: influences of age of acquisition and native language. *Applied Psycholinguist*, 21, 395–423.
- McQueen, J. M., & Huettig, F. (2014). Interference of spoken word recognition through phonological priming from visual objects and printed words. *Attention, Perception, & Psychophysics*, 76(1), 190-200.
- McQueen, J. M., & Viebahn, M. C. (2007). Tracking recognition of spoken words by tracking looks to printed words. *The Quarterly Journal of Experimental Psychology*, 60(5), 661-671.
- Min, C. (2008). The problems and solutions of English immersion education. *Journal of English Education*, 7(1), 109–123.
- Montrul, S. (2011). Multiple interfaces and incomplete acquisition. *Lingua*, 121, 591-604.
- Montrul, S., & Bowles, M. (2009). Back to basics: differential object marking under incomplete acquisition in Spanish heritage speakers. *Bilingualism: Language and Cognition*, 12 (3), 363–383.
- Montrul, S., Foote, R., & Perpiñán, S. (2008). Gender agreement in adult second language learners and Spanish heritage speakers: The effects of age and context of acquisition. *Language Learning*, 58 (3), 503–553.
- Montrul, S., & Ionin, T. (2010). Transfer effects in the interpretation of definite articles by Spanish heritage speakers. *Bilingualism: Language and Cognition*, 13, 449-473.
- Montrul, S., & Slabakova, R. (2003). Competence similarities between native and near-native speakers. *Studies in second language acquisition*, 25(3), 351-398.

- Morgan-Short, K., Sanz, C., Steinhauer, K., & Ullman, M. T. (2010). Second language acquisition of gender agreement in explicit and implicit training conditions: An event-related potential study. *Language Learning*, 60, 154–193.
- Mousavi, S. Y., Low, R., & Sweller, J. (1995). Reducing cognitive load by mixing auditory and visual presentation modes. *Journal of Educational Psychology*, 87, 319–334.
- Müller, G.E., & Pilzecker, A. (1900). Experimental contributions to memory theory. *Zeitschrift für Psychologie Ergänzungsband*. 1: 1–300.
- Muromatsu, K. (2003). Classifiers and the count/mass distinction. *Functional structure (s), form and interpretation: Perspectives from East Asian languages*, 65-96.
- Odlin, T. (1989). *Language transfer: Cross-linguistic influences in language learning*. Cambridge: Cambridge University Press.
- Ojima, S., Nakata, H., & Kakigi, R. (2005). An ERP study of second language learning after childhood: Effects of proficiency. *Journal of Cognitive Neuroscience*, 17, 1212–1228.
- Paivio, A. (1986). *Mental representation: a dual coding approach*. New York: Oxford University.
- Paris, M. C. (1981). *Problèmes de syntaxe et de sémantique en linguistique chinoise* (Vol. 20). (diffusion) Presses universitaires de France.
- Parodi, T., Schwartz, B. D., & Clahsen, H. (2004). On the L2 acquisition of the morphosyntax of German nominals. *Linguistics*, 42, 669–705.
- Paivio, A. (1986). *Mental representation: a dual coding approach*. New York: Oxford University.
- Peeck, J. (1994). Wissenserwerb mit darstellenden Bildern. [Knowledge acquisition with representational illustrations]. In B. Weidenmann (Ed.), *Wissenserwerb mit Bildern. Instruktionale Bilder in Printmedien, Film/Video und Computerprogrammen* (pp. 59–94). Bern: Huber.
- Pelaez, N. J., Boyd, D. D., Rojas, J. B., & Hoover, M. A. (2005). Prevalence of blood circulation misconceptions among prospective elementary teachers. *Advances in Physiology Education*, 29, 172–181.



- Peng, D., & Wang, C. (1997). Basic processing unit of Chinese character recognition: Evidence from stroke number effect and radical number effect [In Chinese]. *Acta Psychologica Sinica*, 29(1), 8–16.
- Penney, C. (1989). Modality effects and the structure of short-term verbal memory. *Memory & Cognition*, 17, 398-422.
- Perani, D., Paulesu, E., Galles, N. S., Dupoux, E., Dehaene, S., Bettinardi, V., Cappa, S. F., Fazio, F., & Mehler, J. (1998). The bilingual brain: Proficiency and age of acquisition of the second language. *Brain*, 121, 1841–1852.
- Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, 36(04), 329-347.
- Plass, J. L., & Jones, L. C. (2005). Multimedia learning in second language acquisition. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 467–488). New York: Cambridge University Press.
- Plass, J. L., Chun, D. M., Mayer, R. E., & Leutner, D. (1998). Supporting visual and verbal learning preferences in a second-language multimedia learning environment. *Journal of Educational Psychology*, 90, 25–36.
- Polio, C. (1994). Non-native speakers' use of nominal classifiers in Mandarin Chinese. *Journal of the Chinese Language Teachers Association*, 29(3): 51-66.
- Pollatsek, A., Drieghe, D., Stockall, L., & de Almeida, R. G. (2010). The interpretation of ambiguous trimorphemic words in sentence context. *Psychonomic Bulletin & Review*, 17(1), 88-94.
- Prévost, P., & White, L. (2000a). Missing surface inflection or impairment in second language acquisition? Evidence from tense and agreement. *Second language research*, 16(2), 103-133.
- Prévost, P., & White, L. (2000b). Accounting for morphological variation in second language acquisition: truncation or missing inflection. *The acquisition of syntax: Studies in comparative developmental linguistics*, 202-235.
- Rayner, K. (1986). Eye movements and the perceptual span in beginning and skilled readers. *Journal of Experimental Child Psychology*, 41, 211–236.

- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124, 372–422.
- Rayner, K. (2009). Eye movements in reading: Models and data. *Journal of Eye Movement Research*, 2(5), 1–10.
- Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition*, 14(3), 191–201.
- Rayner, K., Li, X., Williams, C. C., Cave, K. R., & Well, A. D. (2007). Eye movements during information processing tasks: Individual differences and cultural effects. *Vision research*, 47(21), 2714–2726.
- Rayner, K., Reichle, E. D., & Pollatsek, A. (2005). Eye movement control in reading and the EZ Reader model. *Cognitive processes in eye guidance*, 131–162.
- Rayner, K., Sereno, S. C., Morris, R. K., Schmauder, A. R., & Clifton, C. (1989). Eye movements and on-line language comprehension processes. *Language and Cognitive Processes*, 4, 21–49.
- Rayner, K., Warren, T., Juhasz, B.J., & Liversedge, S.P. (2004). The effect of plausibility on eye movements in reading. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 30(6), 1290–1301.
- Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, 6, 317–327.
- Rebuschat, P., & Williams, J. N. (2012). Implicit and explicit knowledge in second language acquisition. *Applied Psycholinguistics*, 33(04), 829–856.
- Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The EZ Reader model of eye-movement control in reading: Comparisons to other models. *Behavioral and brain sciences*, 26(4), 445–476.
- Roberts, L., Gullberg, M., & Indefrey, P. (2008). Online pronoun resolution in L2 discourse: L1 influence and general learner effects. *Studies in Second Language Acquisition*, 30, 333–357.
- Robinson, P. (2005). Cognitive abilities, chunk-strength, and frequency effects in implicit artificial grammar and incidental L2 learning: Replications of Reber,

- Walkenfeld, & Hernstadt (1991) and Knowlton & Squire (1996) and their relevance for SLA. *Studies in Second Language Acquisition*, 27, 235–268.
- Rossi, S., Gugler, M. F., Friederici, A. D., & Hahne, A. (2006). The impact of proficiency on syntactic second-language processing of German and Italian: Evidence from event-related potentials. *Journal of cognitive neuroscience*, 18(12), 2030-2048.
- Rullmann, H., & You, A. (2006). General number and the semantics and pragmatics of indefinite bare nouns in Mandarin Chinese. *Where semantics meets pragmatics*, 175-196.
- Saalbach, H., & Imai, M. (2005). *Do classifier categories structure our concepts*. Paper presented at the Proceedings of the 26th annual conference of the cognitive science society.
- Saalbach, H., & Imai, M. (2007). Scope of linguistic influence: does a classifier system alter object concepts? *Journal of Experimental Psychology: General*, 136(3), 485.
- Sabourin, L. (2003). Grammatical gender and second language processing: An ERP study. Ph.D. dissertation, Rijksuniversiteit Groningen.
- Sagarra, N., & Ellis, N.C. (2013). From seeing adverbs to seeing morphology. Language experience and adult acquisition of L2 tense. *Studies in Second Language Acquisition*, 35, 261–290.
- Sagarra, N., & Herschensohn, J. (2010). The role of proficiency and working memory in gender and number agreement processing in L1 and L2 Spanish. *Lingua*, 120, 2022–2039.
- Sagarra, N., & Seibert Hanson, A. (2011). Eye-tracking methodology: A user's guide for linguistic research. *Studies in Hispanic and Lusophone Linguistics*, 4(2), 543–555.
- Sanders, L. D., & Neville, H. J. (2003). An ERP study of continuous speech processing II: Segmentation, semantics, and syntax in non-native speakers. *Cognitive Brain Research*, 15, 215–227.
- Sanford, A. J., & Sturt, P. (2002). Depth of processing in language comprehension: Not noticing the evidence. *Trends in cognitive sciences*, 6(9), 382-386.

- Schachter, J. (1990). On the issue of completeness in second language acquisition. *Second Language Research*, 6, 93–124.
- Schachter, J. (1988). Second language acquisition and its relationship to Universal Grammar. *Applied Linguistics*, 9, 219–235.
- Scherag, A., Demuth, L., Rösler, F., Neville, H. J., & Röder, B. (2004). The effects of late acquisition of L2 and the consequences of immigration on L1 for semantic and morpho-syntactic language aspects. *Cognition*, 93, B97–B108.
- Schwartz, B. D., & Sprouse, R. (1994). Word order and nominative case in nonnative language acquisition: a longitudinal study of (L1 Turkish) German interlanguage. *Language acquisition studies in generative grammar*, 31(4), 71-89.
- Schwartz, B., & Sprouse, R. (1996). L2 cognitive states and the full transfer/full access hypothesis. *Second Language Research*, 12, 40–72.
- Schwartz, B., & Sprouse, R. (2010). When syntactic theories evolve: Consequences for L2 acquisition research. *Second language acquisition and linguistic theory*. J. Archibald. Oxford, Blackwell: 156-186.
- Segers, E., Verhoeven, L., & Hulstijn-Hendrikse, N. (2008). Cognitive processes in children's multimedia text learning. *Applied Cognitive Psychology*, 22, 375–387.
- Seidenberg, M. S., Tanenhaus, M. K., Leiman, J. M., & Bienkowski, M. (1982). Automatic access of the meanings of ambiguous words in context: Some limitations of knowledge-based processing. *Cognitive psychology*, 14(4), 489-537.
- Sereno, S. C., Brewer, C. C., & O'Donnell, P. (2003). Context effects in word recognition: Evidence for early interactive processing. *Psychological Science*, 14, 328–333.
- Sereno, S. C., O'Donnell, P. J., & Rayner, K. (2006). Eye movements and lexical ambiguity resolution: Investigating the subordinate bias effect. *Journal of Experimental Psychology: Human Perception and Performance*, 32, 335–350.
- Shanks, D. R., & John, M. F. S. (1994). Characteristics of dissociable human learning systems. *Behavioral and brain sciences*, 17(03), 367-395.

- Sharwood Smith, M. A. (1980). Strategies, language transfer and the simulations of learners' mental operations. *Language Learning*, 29, 345–361.
- Simpson, G. B. (1984). Lexical ambiguity and its role in models of word recognition. *Psychological Bulletin*, 96, 316–340.
- Simpson, G. B., & Burgess, C. (1985). Activation and selection processes in the recognition of ambiguous words. *Journal of Experimental Psychology: Human Perception and Performance*, 11, 28–39.
- Slabakova, R. (2006). Is there a critical period for semantics? *Second Language Research*, 22 (3), 302–338.
- Slabakova, R. (2008). *Meaning in the second language*. Berlin: Walter de Gruyter.
- Slabakova, R. (2013). Adult second language acquisition: A selective overview with a focus on the learner linguistic system. *Linguistic Approaches to Bilingualism*, 3, 48–72.
- Slabakova, R. (2016). *Second language acquisition*. Oxford, UK: Oxford University Press.
- Soja, N. N., Carey, S., & Spelke, E. S. (1991). Ontological categories guide young children's inductions of word meaning: object terms and substance terms. *Cognition*, 38: 179–211.
- Srinivasan, M. (2010). Do classifiers predict differences in cognitive processing? A study of nominal classification in Mandarin Chinese. *Language and Cognition*, 2(02), 177–190.
- Steinhauer, K., White, E. J., & Drury, J. E. (2009). Temporal dynamics of late second language acquisition: Evidence from event-related brain potentials. *Second Language Research*, 25(1), 13–41.
- Su, Y., & Samuels, S. J. (2010). Developmental changes in character-complexity and word-length effects when reading Chinese script. *Reading and Writing*, 23, 1085–1108.
- Sweller, J. (2005). Implications of cognitive load theory for multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 19–30). Cambridge: Cambridge University.

- Sweller, J., Chandler, P., Tierney, P., & Cooper, M. (1990). Cognitive load and selective attention as factors in the structuring of technical material. *Journal of Experimental Psychology: General*, 119, 176–192.
- Sweller, J., Ayres, P., & Kalyuga, S., (2011). *Cognitive load theory*. New York: Springer.
- Swinney, D. A. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of verbal learning and verbal behavior*, 18(6), 645-659.
- Tabbers, H. K., Martens, R. L., & van Merriënboer, J. J. G. (2004). Multimedia instructions and cognitive load theory: effects of modality and cueing. *British Journal of Educational Psychology*, 74, 71–81.
- Tai, J. H. (1992). Variation in classifier systems across Chinese dialects: towards a cognition-based semantic approach. *Chinese Language and Linguistics*, 1, 587-608.
- Tai, J. H-Y. (1994). Chinese classifier systems and human categorization. In *In honor of William S. Y. Wang: Interdisciplinary studies on language and language change*. Ed. M.Y. Chen and O. Tzeng, 479-494. Taipei: Pyramid Press.
- Tai, J., & Wang, L. (1990). A Semantic Study of the Classifier Tiao. *Journal of the Chinese Language Teachers Association*, 25(1), 35-56.
- Tanenhaus, M. K., Leiman, J. M., & Seidenberg, M. S. (1979). Evidence for multiple stages in the processing of ambiguous words in syntactic contexts. *Journal of verbal learning and verbal behavior*, 18(4), 427-440.
- Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268(5217), 1632-1634.
- Tanenhaus, M.K., & Trueswell, J.C. (2006). Eye movements and spoken language comprehension. In M. Traxler & M. Gernsbacher (Eds.), *Handbook of psycholinguistics* (2nd ed.) (pp. 863–900). New York, NY: Elsevier.
- Tang, C. C. J. (1990). Chinese phrase structure and extended X'-theory. PhD thesis, Cornell University.
- Tang, C. C. J. (1996). *Ta mai-le bi shi-zhi* and Chinese phrase structure, *BIHP* 66.

- Tien, Y. M., Tzeng, O. J. L., & Hung, D. L. (2002). Semantic and cognitive basis of Chinese classifiers: A functional approach. *Language and Linguistics*, 3(1), 101–132.
- Tindall-Ford, S., Chandler, P., & Sweller, J. (1997). When two sensory modes are better than one. *Journal of Experimental Psychology: Applied*, 3, 257–287.
- Tinker, M. A. (1946). The study of eye movements in reading. *Psychological Bulletin*, 43(2), 93.
- Tokowicz, N., & MacWhinney, B. (2005). Implicit and explicit measures of sensitivity to violations in second language grammar. *Studies in Second Language Acquisition*, 27, 173 – 204.
- Tolentino, L. C., & Tokowicz, N. (2011). Across languages, space, and time. A review of the role of cross-language similarity in L2 (morpho)syntactic processing as revealed by fMRI AND ERP methods. *Studies in Second Language Acquisition*, 33, 91-125.
- Tolentino, L. C., & Tokowicz, N. (2014). Cross-language similarity modulates effectiveness of second language grammar instruction. *Language Learning*, 64(2), 279-309.
- Townsend, D. J., & Bever, T. G. (2001). *Sentence comprehension: The integration of habits and rules* (Vol. 1950). Cambridge, MA: MIT Press.
- Trenkic, D. (2002). Form–meaning connections in the acquisition of English articles. In S. Foster-Cohen, T. Ruthenberg & M. L. Poschen (eds.), *EUROSLA Yearbook* (vol. 2), pp. 115–133. Amsterdam: John Benjamins.
- Trenkic, D. (2007). Variability in L2 article production: Beyond the representational deficit vs. processing constraints debate. *Second Language Research*, 23, 289–327.
- Trenkic, D., Mirkovic, J., & Altmann, G. T. M. (2014). Real-time grammar processing by native and non-native speakers: constructions unique to the second language. *Bilingualism: Language and Cognition*, 17(2), 237-257.
- Trenkic, D., & Pongpiroj, N. (2013). Referent salience affects second language article use. *Bilingualism: Language and Cognition*, 16, 152–166.

- Tsimpli, I. M., & Dimitrakopoulou, M. (2007). The interpretability hypothesis: Evidence from wh-interrogatives in second language acquisition. *Second Language Research*, 23, 215–242
- T'sou, B. K. (1976). The structure of nominal classifier systems. *Oceanic linguistics special publications*, (13), 1215-1247.
- Ullman, M. T. (2001). The neural basis of lexicon and grammar in first and second language: The declarative/procedural model. *Bilingualism: Language and cognition*, 4(02), 105-122.
- Ullman, M. T. (2005) A cognitive neuroscience perspective on second language acquisition: the declarative/procedural model. In *Mind and Context in Adult Second Language Acquisition: Methods, Theory, and Practice* (Sanz, C., ed.), pp. 141–178, Georgetown University Press
- Underwood, B. J. (1948). Spontaneous recovery' of verbal associations. *Journal of Experimental Psychology*. 38: 429–439.
- van Hell, J. G., & Tokowicz, N. (2010). Event-related brain potentials and second language learning: Syntactic processing in late L2 learners at different L2 proficiency levels. *Second Language Research*, 26, 43 – 74.
- Wang, H. C., Pomplun, M., Chen, M., Ko, H., & Rayner, K. (2010). Estimating the effect of word predictability on eye movements in Chinese reading using latent semantic analysis and transitional probability. *The Quarterly Journal of Experimental Psychology*, 63(7), 1374-1386.
- Wang, M. (2011). Learning a second language. In R. E. Mayer, & P. A. Alexander (Eds.), *Handbook of research on learning and instruction* (pp. 127–147). New York: Routledge.
- Wang, S., Chen, H.-C., Yang, J., & Mo, L. (2008). Immediacy of integration in discourse comprehension: Evidence from Chinese readers' eye movements. *Language and Cognitive Processes*, 23, 241–257.
- Warren, T., White, S. J., & Reichle, E. D. (2009). Investigating the causes of wrap-up effects: Evidence from eye movements and E-Z Reader. *Cognition*, 111, 132-137.



- Wen, Z., Miyao, M., Takeda, A., Chu, W., & Schwartz, B. D. (2010). Proficiency effects and distance effects in non-native processing of English number agreement. In K. Franich, K. M. Iserman, & L. L. Keil (Eds.), *Proceedings of the 34th Boston University Conference on Language Development* (pp. 445–456). Somerville, MA: Cascadilla Press.
- White, L. (1989). *Universal Grammar and second language acquisition*. Amsterdam, The Netherlands: John Benjamins Publishing.
- White, L. (2003). *Second language acquisition and universal grammar*. Cambridge, UK: Cambridge University Press.
- White, L. (2011). Second language acquisition at the interfaces. *Lingua*, 121, 577–590.
- Wickham, H. (2009). *ggplot2: elegant graphics for data analysis*. Springer Science & Business Media.
- Williams, J. N. (2005). Learning without awareness. *Studies in Second Language Acquisition*, 27, 269–304.
- Williams, J. N. (2009). Implicit learning in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *The new handbook of second language acquisition*. Chapter 14, 319-353. Bingley, UK: Emerald.
- Williams, J. N., & Lovatt, P. (2005). Phonological memory and rule learning. *Language Learning*, 55, 177–233.
- Yan, G., Tian, H., Bai, X., & Rayner, K. (2006). The effect of word and character frequency on the eye movements of Chinese readers. *British Journal of Psychology*, 97, 259–268.
- Yan, M., Zhou, W., Shu, H., Yusupu, R., Miao, D., Krügel, A., & Kliegl, R. (2014). Eye movements guided by morphological structure: *Evidence from the Uighur language*. *Cognition*, 132(2), 181-215.
- Yang, H. M., & McConkie, G.W. (1999). Reading Chinese: Some basic eye- movement characteristics. In H.-C. Chen (Ed.), *Reading Chinese script: A cognitive analysis* (pp. 207–222). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

- Yang, J., Wang, S., Tong, X., & Rayner, K. (2012). Semantic and plausibility effects on preview benefit during eye fixations in Chinese reading. *Reading and Writing*, 25(5), 1031-1052.
- Yao, P., & Chen, B. (2016). Cross-linguistic differences affect late Chinese-English learners on-line processing of English tense and aspect. *International Journal of Bilingualism*, 1367006915624248.
- Yi, B. (2011). Afterthoughts on Chinese Classifiers and Count Nouns. *Plurality in Classifier languages (2011)* Kim (ed.), Hankookmunhwasa, 265-282.
- Ying, H., Chen, G., Song, Z., Shao, W., & Guo, Y. (1983). 4-7 Sui Ertong Zhangwo Liangci De Tedian [Characteristics of 4-to-7-year-olds in Mastering Classifiers]. *Information on Psychological Sciences*, 26, 24-32.
- Yu, B., & Cao, H. (1992). A new exploration on the effect of stroke-number in the identification of Chinese characters [In Chinese]. *Acta Psychologica Sinica*, 2, 120-126.
- Zhang, J. J., & Liu, Y. H. (2009). The lexical access of individual classifiers in language production and comprehension [In Chinese]. *Acta Psychologica Sinica*, 41(7), 580-593.
- Zhang, N. N. (2010). The constituency of classifier constructions in Mandarin Chinese. *Taiwan Journal of Linguistics*, 9(1), 1-50.
- Zhang, N. N. (2012). Countability and numeral classifiers in Mandarin Chinese. *A Cross Linguistic Exploration of the Count Mass Distinction*, 220-237.
- Zhang, Y., Shu, H., Zhang, H., & Zhou, X. (2002). The Processing of Ambiguous Phrases in Referential Discourse Context in Chinese (Article written in Chinese). *Acta Psychologica Sinica*, 34(2), 126-134.
- Zuur, A. F., Ieno, E. N., Walker, N. J., Saveliev, A. A., & Smith, G. M. (2009). Zero-truncated and zero-inflated models for count data. *In Mixed effects models and extensions in ecology with R* (pp. 261-293). Springer New York.